THE INTERACTION BETWEEN OBJECTIVE ANALYSIS AND INITIALIZATION

*Proceedings of the (Fourteenth) Stanstead Seminar held at Bishop’s University, Lennoxville, Québec Canada*

R. Daley, J. Derome, and D. Williamson, Chairmen
*July 12-16, 1982*

PUBLICATION IN METEOROLOGY 127 December 1982
D. Williamson, Editor

Dépôt légal — Quatrième Trimestre 1982
Bibliothèque nationale du Québec
ABSTRACT

A collection of summaries of lectures presented at the Fourteenth Stanstead Seminar on "the interaction between objective analysis and initialization." The lectures both reviewed the fundamentals of commonly used analysis and initialization techniques and presented recent work, with emphasis on current problems.

RESUMÉ

Une collection des sommaires des conférences présentées au quatorzième colloque de Stanstead. Les présentations passèrent en revue les méthodes courantes d'analyse et d'initialisation et décrivent certains travaux récents dans ces domaines en mettant l'accent sur les difficultés qui restent à surmonter.
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>v</td>
</tr>
<tr>
<td>List of Participants</td>
<td>vii</td>
</tr>
<tr>
<td>Principal acronyms</td>
<td>viii</td>
</tr>
<tr>
<td>Optimal interpolation: Basic formulation and characteristics</td>
<td>1</td>
</tr>
<tr>
<td>by R. D. McPherson</td>
<td></td>
</tr>
<tr>
<td>Optimal interpolation: Practical aspects of operational application</td>
<td>7</td>
</tr>
<tr>
<td>by R. D. McPherson</td>
<td></td>
</tr>
<tr>
<td>Fundamentals of nonlinear normal mode initialization</td>
<td>13</td>
</tr>
<tr>
<td>by B. Machenhauer</td>
<td></td>
</tr>
<tr>
<td>Characteristic structures of atmospheric flow and temperature as</td>
<td>44</td>
</tr>
<tr>
<td>seen from FGGE analyses</td>
<td></td>
</tr>
<tr>
<td>by F. Baer</td>
<td></td>
</tr>
<tr>
<td>Initialization experiments with the NMC spectral model</td>
<td>52</td>
</tr>
<tr>
<td>by B. A. Ballish</td>
<td></td>
</tr>
<tr>
<td>A comparison of two static initialization methods in data assimila-</td>
<td>57</td>
</tr>
<tr>
<td>tion</td>
<td></td>
</tr>
<tr>
<td>by E. H. Barker</td>
<td></td>
</tr>
<tr>
<td>The global weather experiment - results and proposals for future</td>
<td>62</td>
</tr>
<tr>
<td>observing systems</td>
<td></td>
</tr>
<tr>
<td>by L. Bengtsson</td>
<td></td>
</tr>
<tr>
<td>Convergence properties of the Machenhauer initialization scheme</td>
<td>77</td>
</tr>
<tr>
<td>by R. Errico</td>
<td></td>
</tr>
<tr>
<td>Sequential estimation, data assimilation and initialization</td>
<td>83</td>
</tr>
<tr>
<td>by M. Ghil, S. E. Cohn and A. Dalcher</td>
<td></td>
</tr>
<tr>
<td>Operational forecasting at ECMWF</td>
<td>98</td>
</tr>
<tr>
<td>by A. Hollingsworth</td>
<td></td>
</tr>
<tr>
<td>Objective analysis of streamfunction and velocity potential for</td>
<td>100</td>
</tr>
<tr>
<td>analysis of tropical wind fields</td>
<td></td>
</tr>
<tr>
<td>by P. R. Julian</td>
<td></td>
</tr>
<tr>
<td>Application of the bounded derivative method to atmospheric modeling</td>
<td>106</td>
</tr>
<tr>
<td>by A. Kasahara</td>
<td></td>
</tr>
<tr>
<td>Numerical weather prediction experiments over the tropics using a</td>
<td>113</td>
</tr>
<tr>
<td>global spectral model</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Adjustment of VAS geopotential analysis by using precedent RAOB</td>
<td>115</td>
</tr>
<tr>
<td>analyses and quasi-geostrophic constraints</td>
<td></td>
</tr>
<tr>
<td>by J. Lewis</td>
<td></td>
</tr>
<tr>
<td>On direct data assimilation into a forecast model</td>
<td>126</td>
</tr>
<tr>
<td>by A. C. Lorenc</td>
<td></td>
</tr>
<tr>
<td>On the completeness of multivariate optimum interpolation</td>
<td>132</td>
</tr>
<tr>
<td>by N. A. Phillips</td>
<td></td>
</tr>
<tr>
<td>Kalman filtering in a two-point atmosphere</td>
<td>134</td>
</tr>
<tr>
<td>by N. A. Phillips</td>
<td></td>
</tr>
<tr>
<td>Normal mode initialization in the ANMRC data assimilation scheme</td>
<td>137</td>
</tr>
<tr>
<td>by K. Puri</td>
<td></td>
</tr>
<tr>
<td>Variational methods in meteorology</td>
<td>143</td>
</tr>
<tr>
<td>by Y. Sasaki</td>
<td></td>
</tr>
<tr>
<td>Four-dimensional assimilation at GFDL using FGGE data</td>
<td>150</td>
</tr>
<tr>
<td>by W. Stern</td>
<td></td>
</tr>
<tr>
<td>Variational normal mode initialization for the ECMWF gridpoint model</td>
<td>160</td>
</tr>
<tr>
<td>by C. Temperton</td>
<td></td>
</tr>
<tr>
<td>On the existence of solutions to Machenhauer's nonlinear normal</td>
<td>165</td>
</tr>
<tr>
<td>mode initialization</td>
<td></td>
</tr>
<tr>
<td>by L. Thaning</td>
<td></td>
</tr>
<tr>
<td>Experiments with variational normal mode initialization</td>
<td>171</td>
</tr>
<tr>
<td>by J. Tribbia</td>
<td></td>
</tr>
<tr>
<td>Variational methods in simultaneous optimum interpolation and</td>
<td>178</td>
</tr>
<tr>
<td>initialization</td>
<td></td>
</tr>
<tr>
<td>by G. Wahba</td>
<td></td>
</tr>
<tr>
<td>A unified analysis-initialization technique</td>
<td>186</td>
</tr>
<tr>
<td>by D. Williamson and R. Daley</td>
<td></td>
</tr>
<tr>
<td>What I learned at the 1982 Stanstead Seminar</td>
<td>192</td>
</tr>
<tr>
<td>by C. Farquarson (transmitted by P. Merilees)</td>
<td></td>
</tr>
</tbody>
</table>
There is mounting evidence that the specification of the initial atmospheric state is as much to blame for errors in numerical weather forecasts as the mathematical formulation of the model or the parameterization of physical processes. The complete forecast system is comprised of four components—observations, analysis, initialization, and numerical model—with three corresponding interfaces. Analysis schemes produce an estimate of atmospheric state variables distributed uniformly in space and time from the irregular and incomplete set of observations. Initialization schemes modify the analyses to make them acceptable as initial states for models so that the model forecasts do not contain spurious waves which obscure or even interfere with the development of meteorologically important features.

To a large extent, these four components have been developed independently by specialists in each area with limited but not complete understanding of the principles involved in all the others. But there are notable instances of collaboration among specialists in problems related to the interface between observations and analysis, and the interface between initialization and models. For example, statistical objective analysis or optimal interpolation takes into account the error characteristics and distribution of various observations to produce an estimate of the analysis error. Dynamic initialization and nonlinear normal mode initialization both use the forecast model itself as a basis of determining the necessary changes to the analysis to make the data acceptable to that model. The interface between analysis and initialization has received less attention. Although four-dimensional data assimilation is an attempt to bridge this gap, a unifying principle is still lacking.
The (Fourteenth) Stanstead Seminar on "The Interaction Between Objective Analysis and Initialization" was organized to bring together specialists from both areas to stimulate cross-fertilization. Speakers were invited to review the fundamentals of commonly used analysis and initialization methods for the benefit of specialists in other areas and to present recent work with emphasis on current problems and difficulties.

This publication includes the speakers' summaries of papers presented at this Fourteenth Stanstead Seminar held in the summer of 1982 at Bishop's University in Lennoxville, Quebec. The program consisted of twenty-six papers presented over a five-day period. Two speakers were asked to review the basics of optimal interpolation and normal mode initialization. Their contributions have been placed at the front of the Proceedings. The others, with a heavier emphasis on current research, have been placed in alphabetical order. This year's seminar benefited from the participation of scientists from leading American, European, Australian, and Canadian groups involved in numerical weather forecasting as well as from the participation of individuals from various universities active in this field. We would like to take this opportunity to thank all those who participated in this year's seminar and who contributed to its success.

The seminar was supported by McGill's Faculty of Graduate Studies and Research, by the National Center for Atmospheric Research, and through the participation of many people from the Atmospheric Environment Service of Canada.

R. Daley
J. Derome
D. Williamson

December 1982
LIST OF PARTICIPANTS

J. Alpert, NASA
R. Asselin, AES
F. Baer, U. Maryland
B. Ballish, NMC
E. Barker, NEPRF
P. Bartello, McGill
C. Beaudoin, McGill
M. Béland, AES
L. Bengtsson, ECMWF
R. Benoit, AES
P. Betout, Météo Fr.
W. Blumen, U. Colorado
G. Boer, AES
W. Bonner, NMC
C. Chouinard, AES
S. Cohn, Courant Inst.
A. Dalcher, GLAS
R. Daley, NCAR
Y. Delage, AES
J. Derome, McGill
C. Dey, NMC
B. Dugas, McGill
R. Errico, NCAR
P. Gauthier, McGill
M. Ghil, Courant Inst.
J. Gillespie, AES
C. Girard, AES
S. Gravel, McGill
A. Hollingsworth, ECMWF
P. Julian, NCAR
A. Kasahara, NCAR
R. Kistler, NMC
T. Kitade, FSU
P. Koclas, AES
T. Krishnamurti, FSU
S. Lambert, AES
R. Laprise, AES
F.-X. Le Dimet, U. Clermont
J. Lewis, NOAA/NESS
A. Lorenc, UKMO
B. Machenhauer, U. Copenhagen
J. Mailhot, McGill
N. McFarlane, AES
C. McLandress, McGill
R. McPherson, NMC
P. Merilees, AES
H. Mitchell, AES
R. Pasch, FSU
M. Pedder, U. Reading
N. Phillips, NMC
K. Puri, ANMRC
P. Rasch, NCAR
S. Ray, McGill
H. Ritchie, AES
I. Rutherford, AES
D. Salstein, Atm. and Env. Res.
N. Sargent, AES
Y. Sasaki, U. Oklahoma
T. Schlatter, NOAA
D. Shantz, AES
A. Staniforth, AES
W. Stern, GFDL
H. Sundqvist, U. Stockholm
O. Talagrand, LMD (Paris)
C. Temperton, UKMO
L. Thaning, U. Stockholm
J. Tribbia, NCAR
A. Van Tuyl, U. Wisconsin
R. K. Vupputuri, AES
C. Wahba, U. Wisconsin
T. Warn, McGill
D. Williamson, NCAR
E. Yakimiw, AES
C. Yang, AFGL
T.-W. Yu, NMC
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>Atmospheric Environment Service</td>
</tr>
<tr>
<td>AFGL</td>
<td>Air Force Geophysics Laboratory</td>
</tr>
<tr>
<td>ANMRC</td>
<td>Australian Numerical Meteorological Research Centre</td>
</tr>
<tr>
<td>ECMWF</td>
<td>European Centre for Medium Range Weather Forecasts</td>
</tr>
<tr>
<td>GLAS</td>
<td>Goddard Laboratory for Atmospheric Science</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NCAR</td>
<td>National Center for Atmospheric Research</td>
</tr>
<tr>
<td>NEPRF</td>
<td>Naval Environmental Prediction Research Facility</td>
</tr>
<tr>
<td>NMC</td>
<td>National Meteorological Center, NOAA</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanographic and Atmospheric Administration</td>
</tr>
<tr>
<td>NOAA/NESS</td>
<td>National Earth Satellite Service</td>
</tr>
<tr>
<td>UKMO</td>
<td>United Kingdom Meteorological Office</td>
</tr>
</tbody>
</table>
This paper presents a summary of the basic formulation of the method of optimum interpolation, within the very simple context of performing an analysis at one point using only three pieces of information. Some of the method's characteristics are then illustrated by means of a series of simple analysis problems.

We assume for this development that we have one observation of geopotential \( H_0 \), and one of wind which is treated in the form of eastward \( U_0 \) and northward \( V_0 \) components, for a total of three pieces of information. The analyzed values of these parameters \( H_a, U_a, V_a \) at some point in the vicinity of the observation may be written as a "guess" value of the parameter \( H_g, U_g, V_g \), usually from a forecast, plus a correction formed by linear combinations involving the observations.

\[
egin{align*}
H_a &= H_g + a_1(H_0-H_g)_1 + a_2(U_0-U_g)_2 + a_3(V_0-V_g)_3 \\
U_a &= U_g + b_1(H_0-H_g)_1 + b_2(U_0-U_g)_2 + b_3(V_0-V_g)_3 \\
V_a &= V_g + c_1(H_0-H_g)_1 + c_2(U_0-U_g)_2 + c_3(V_0-V_g)_3
\end{align*}
\]

where \((H_0-H_g)_1\), etc. are the differences, or "residuals", between the data at an observation location and the appropriate guess value interpolated to that location, and the \( a_1, b_1, c_1 \) are coefficients which determine the weight each datum receives in the analysis, relative to the guess value. The analysis consists of determining the coefficients.

Note that observations of both geopotential and wind are used in the analyses of both variables: this type of analysis is said to be "multivariate".
In contrast, an analysis which uses only observations of the variable being analyzed is said to be "univariate".

In order to calculate the coefficients \((a_i, b_i, c_i)\), we begin by subtracting the true values \((H_T, U_T, V_T)\) from both sides of eqns. (1), changing the signs, and introducing the following definitions:

\[
\begin{align*}
   f_a &= F_T - F_H: \text{ true analysis error in the variable } F(H, U, \text{ or } V); \\
   f_g &= F_T - F_g: \text{ true guess error in } F; \\
   F_o &= F_T + \varepsilon_f: \varepsilon_f = \text{ error in observation of the variable } F.
\end{align*}
\]

We note that the "observed residuals" \(F_o - F_g\), may be rewritten as

\[
F_o - F_g = F_T - F_g + \varepsilon_f = f_g + \varepsilon_f
\]

so that the observed residual may be expressed as the sum of the error in the guess plus the error in the observation. Eqns. (1) may then be recast in terms of these error quantities:

\[
\begin{align*}
   h_a &= h_g - [a_1(h_g + \varepsilon_h)_1 + a_2(u_g + \varepsilon_u)_2 + a_3(v_g + \varepsilon_v)_3] \\
   u_a &= u_g - [b_1(h_g + \varepsilon_h)_1 + b_2(u_g + \varepsilon_u)_2 + b_3(v_g + \varepsilon_v)_3] \\
   v_a &= v_g - [c_1(h_g + \varepsilon_h)_1 + c_2(u_g + \varepsilon_u)_2 + c_3(v_g + \varepsilon_v)_3]
\end{align*}
\] (2)

Optimum interpolation determines the coefficients \((a_i, b_i, c_i)\) such that the mean-square analysis errors \((h_a^2, u_a^2, v_a^2)\) are minimized.

\[
\frac{\partial}{\partial a_i}(h_a^2) = \frac{\partial}{\partial b_i}(u_a^2) = \frac{\partial}{\partial c_i}(v_a^2) = 0
\] (3)

The minimization process results in sets of three linear equations, one set for each variable: for geopotential,

\[
\begin{align*}
   (h_1 h_1 + \varepsilon_h) a_1 + (h_1 u_2) a_2 + (h_1 v_3) a_3 &= h_1 h_g \\
   (u_2 h_1) a_1 + (u_2 u_2 + \varepsilon_u) a_2 + (u_2 v_3) a_3 &= u_2 h_g \\
   (v_3 h_1) a_1 + (v_3 u_2) a_2 + (v_3 v_3 + \varepsilon_v) a_3 &= v_3 h_g
\end{align*}
\] (4)
The parenthetical quantities are various covariances of guess, or forecast, errors. Note that the subscript g has been deleted from the forecast error at an observation location, but retained in the forecast error at the analysis point. For example, the term \( h_1 h_g \) denotes the covariance between the forecast geopotential error at observation location \( h_1 \) and its counterpart at the analysis point \( h_g \). Quantities on the main diagonal such as \( h_1 h_1 \), etc., are recognized as the variances of the forecast errors \( \sigma_h^2 \), \( \sigma_u^2 \), \( \sigma_v^2 \). The off-diagonal quantities are seen to be cross-covariances between forecast errors of geopotential and those of wind.

It should be noted that terms such as \( h_i \epsilon_j \) and \( \epsilon_h \epsilon_v \) have been assumed to vanish.

For the u- and v-components, similar sets of equations may be obtained, in which the \( a_i \) are replaced by \( b_i \) and \( c_i \) and the \( h_g \) in the covariances on the right-hand side is replaced by \( u_g \) and \( v_g \), respectively. So long as the same set of observations is used for the analysis of geopotential and the wind components, the forecast error covariance matrix on the left-hand side of eqns. (4) is common to all three sets. The order of the three sets is determined by the number of observations used in the analysis.

If the forecast error covariance matrix and the right-hand side vectors can be specified, the matrix can be inverted and the three sets solved for the unknowns \( a_i \), \( b_i \), \( c_i \). The analyzed values at the analysis point may then be calculated from eqns. (1). It may also be shown that the minimized analysis error variance may be determined from

\[
\begin{align*}
\bar{h}_a^2 &= \sigma_h^2 - a_1 h_1 h_g - a_2 u_2 h_g - a_3 v_3 h_g \\
\bar{u}_a^2 &= \sigma_u^2 - b_1 h_1 u_g - b_2 u_2 u_g - b_3 v_3 u_g \\
\bar{v}_a^2 &= \sigma_v^2 - c_1 h_1 v_g - c_2 (u_2 v_g) - c_3 v_3 v_g.
\end{align*}
\]
It is computationally convenient to model the forecast error covariance matrix by an analytic, differentiable function which approximates actual forecast error covariances. In the NMC system, the height-height error covariance is specified by

$$\overline{h_i h_j} = \{\sigma(h)_i \sigma(h)_j\} \{\exp[-K_h(s_i - s_j)^2]\} \{1 + K_p \ln^2(p_i/p_j)\}^{-1}$$

(6)

where $s_i - s_j$ is the horizontal separation between points (i) and (j), and $p_i - p_j$ is their vertical separation. This expression is in the form of a triple product of a variance, a horizontal correlation, and a vertical correlation. The forecast error cross-covariances in eqns. (4) are determined by assuming that the errors in geopotential, temperature, and wind are related hydrostatically and geostrophically. Under these assumptions, all forecast error covariances can be calculated by differentiation of eqn. (6). Table 1 gives the differential relationship between $\overline{h_i h_j}$ and all other covariances.

Table 1. Covariance of the row variable with the column variable in terms of the geopotential autocovariance $\overline{h h}$, assuming that height, temperature, and wind residuals are related through the geostrophic and hydrostatic equations.

<table>
<thead>
<tr>
<th></th>
<th>h</th>
<th>t</th>
<th>u</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>$\overline{h h}$</td>
<td>$- \frac{\partial \overline{h h}}{\partial z}$</td>
<td>$- \frac{\partial \overline{h h}}{\partial y}$</td>
<td>$\frac{\partial \overline{h h}}{\partial x}$</td>
</tr>
<tr>
<td>t</td>
<td>$- \frac{\partial \overline{h h}}{\partial \zeta}$</td>
<td>$(\frac{\partial}{\partial \zeta})^2 \overline{h h}$</td>
<td>$\frac{\partial^2 \overline{h h}}{\partial \zeta \partial y}$</td>
<td>$- \frac{\partial^2 \overline{h h}}{\partial \zeta \partial x}$</td>
</tr>
<tr>
<td>u</td>
<td>$- \frac{\partial \overline{h h}}{\partial \theta}$</td>
<td>$\frac{\partial^2 \overline{h h}}{\partial \theta \partial z}$</td>
<td>$\frac{\partial^2 \overline{h h}}{\partial \theta \partial y}$</td>
<td>$- \frac{\partial^2 \overline{h h}}{\partial \theta \partial x}$</td>
</tr>
<tr>
<td>v</td>
<td>$- \frac{\partial \overline{h h}}{\partial \xi}$</td>
<td>$(\frac{\partial}{\partial \xi})^2 \overline{h h}$</td>
<td>$\frac{\partial^2 \overline{h h}}{\partial \xi \partial z}$</td>
<td>$- \frac{\partial^2 \overline{h h}}{\partial \xi \partial x}$</td>
</tr>
</tbody>
</table>
The main diagonal terms of eqns. (4) contain forecast error and observational error variances which must be specified. With respect to the former, it should be noted that the assumption of a hydrostatic and geostrophic relationship between forecast errors implies a relationship between the variances of the errors as well. It can be shown that

\[ \sigma_u^2 = \sigma_n^2 \left( \frac{2Kg^2}{f^2} \right), \]  

and

\[ \sigma_t^2 = \sigma_n^2 \left( \frac{2Kg^2}{R^2} \right), \]  

where \( \sigma_t \) is the forecast temperature error variance.

In an analysis/forecast data assimilation cycle, it is necessary to predict the forecast geopotential error variance valid at the next analysis time, usually a few hours in advance. For the NMC 6h cycle, this is accomplished by assuming that at the next analysis time (denoted by superscript \( n+1 \)) is related to the analysis error variance at time \( n \):

\[ \sigma_{h}^{n+1} = \left( \sqrt{\sigma} \right)^n + D \]  

where \( D \) is an estimate of the forecast error growth rate determined from verification statistics. In the NMC system, \( D \) is a function of variable and pressure level, but not horizontal position.

The observational error variances (\( \sigma_{h}^2, \sigma_{u}^2, \sigma_{v}^2 \)) in the covariance matrix must be pre-specified. Typically, this is done by classes of observations; that is, radiosondes are assigned one error variance, satellite data another, aircraft another, etc. Current values
in use at NMC were adopted from the European Centre for Medium-Range Weather Forecasting and may be found in Bengtsson (1981).*

Some of the characteristics of the method may be listed as follows:

- Observations are weighted in proportion to the ratio of the accuracy of the data to the accuracy of the forecast in the vicinity of the observation: the more accurate the data relative to the forecast the more weight it receives in the analysis;
- All else being equal, observations with smaller error variances receive more weight than those with larger ones;
- The effect on non-independence of observations is recognized by reducing the weight such observations receive in univariate analyses;
- In the analysis of geopotential, a wind observation added to a geopotential observation is more beneficial than a geopotential observation by itself.
- Observations with random errors receive more weight in a univariate analysis than those with correlated errors; but it is better to have correlated than random errors in an analysis where observations of one variable are being used to analyze another variable related to the gradient of the first.

Illustrations of these characteristics may be found in NMC Office Note 265.

* Bengtsson, L., 1981: "Data analysis, initialization, and data assimilation."

Operational experience with optimum interpolation at NMC and elsewhere has stimulated close examination of the characteristics of the method. Some of these questions have arisen out of the basic formulation of the method; for example, how sensitive is the resulting analysis to variations in the theoretical forecast error covariance model? Others result from compromises necessary to implement optimum interpolation on existing computers for execution within operational deadlines. For example, computer limitations require restricting the amount of data influencing the analysis at any point; one can then enquire what the implications of this are, and what is the proper way to select observations to be used.

At NMC, examination of these aspects of optimum interpolation came about as a result of questions concerning the method's ability to resolve relatively small-scale features in atmospheric flow patterns. In the 6-hour analysis/forecast cycle which constitutes the NMC Global Data Assimilation System, occasions have been frequently noted in which rapid cyclogenesis is not adequately represented. Usually, this is manifest by the 6h predictions of the cyclogenesis being too slow and of insufficient intensity. The corrections which the optimum interpolation analysis makes in such cases tend to be localized, of relatively small scale, but sometimes of considerable amplitude. Maps of these correction fields display "bullseyes", with horizontal dimensions ranging from about 10 degrees latitude in diameter to about 40 degrees with amplitudes (in the geopotential field) of more than 100m in mid-troposphere, in extreme cases.

Early in the operational life of the optimum interpolation system at NMC, it was noted that features on the small end of the length scale range were not analyzed as accurately as those with larger dimensions; that is,
the analyses in such cases did not reflect the data as faithfully as might be desired. In particular, small-scale features were not analyzed as small enough in horizontal dimensions, or with enough intensity.

To explore the factors influencing this apparent scale limitation, a series of one-dimensional analysis simulation experiments was performed. The characteristics of the analysis model are outlined as follows:

- Analysis grid: 1-dimensional, along latitude 45, with analysis points at 2° intervals
- Data: True correction field specified analytically as a function of longitude:

\[
    h(\lambda) = A \cos \frac{2\pi \lambda}{L}
\]

\[
    v(\lambda) = \frac{g}{f \cos \phi} \frac{\partial h}{\partial \lambda}
\]

Observations composed of true field plus random error with zero mean and standard deviation \(E(h), E(V)\).

32 equally spaced observation points at 2.5° intervals 64 observations total

- Statistics: First guess error covariance

\[
    \overline{h_i h_j} = \sigma_h^2 \ e^{-ks^2}, \ s = \text{separation distance between } i \text{ and } j
\]

\[
    \overline{h_i v_j} = \frac{g}{f \cos \phi} \frac{\partial (h_i h_j)}{\partial \lambda}
\]

\[
    \overline{v_i v_j} = \frac{g^2}{f^2 r^4 \cos^2 \phi} \frac{\partial^2}{\partial \lambda^2} (h_i h_j)
\]

First guess error variance

\[
    \sigma_h = 40 \text{ m}
\]

\[
    \sigma_v = \left(\frac{g}{f} \sqrt{2k}\right) \sigma_h
\]

Observational error

\[
    E(h) = 20 \text{ m (in most experiments)}
\]

\[
    E(V) = \left(\frac{g}{f} \sqrt{2k}\right) E(h)
\]
To illustrate the problem of scale limitation, the simulated true correction field was given horizontal dimensions varying from 40° latitude to 10° latitude with \( k=1 \times 10^{-6} \text{km}^{-2} \). Figure 1 shows the response of the analysis to a feature 25° latitude from zero value to zero value. The solid lines depict the "truth" - that is, the analytic field - and the dotted lines represent the analysis. Simulated observations are shown by crosses. In this experiment, all 32 geopotential observations and 32 wind observations were used in the analysis at every point. It will be noted that the analyzed geopotential is a close approximation of the truth, and that the random noise in the data has been eliminated: the minimum value in the true field, 100m, is analyzed as 91m, and the analysis is agreeably smooth.

By comparison, Figure 2 shows the same depiction but with the dimensions of the true correction field reduced to 10° latitude. The minimum analyzed value is now only 53m, and the analyzed dimensions are larger - the wave length is about 22° latitude.

The effort to investigate this behavior included simulation experiments on:

- multivariate vs. univariate analysis;
- variations in the forecast error covariance model
  \( (k=1,2,4 \times 10^{-6} \text{km}^{-2}) \);
- variations in the number of observations used
  \( (5, 8, 10, 20, 64) \);
- variations in the observational standard deviations
  \( (\sigma = 20, 10, 5, 2\text{m}) \);
- variations in data selection method
  (closest vs. most highly correlated).
It was concluded that:

- The scale limitation in optimum interpolation is controlled mostly by the shape of the covariance model, and to some extent by the quality of the data.
- For the covariance model in use at NMC (Gaussian, $k=10^{-6}$ km$^{-2}$), features in the correction field of scale less than about 20° latitude will not be represented faithfully.
- For data with random errors, it is important to use enough data to reduce the effect of errors, but it does not appear necessary to use all of the data.
- Data selection should result in the same set of observations being used in both mass and motion analyses.

A more detailed exposition of these results may be found in NMC Office Note 266.
Figure 1. One dimensional analysis simulation, with the length of the true correction field (forecast minus observed) specified as 25 degrees latitude. Solid lines represent the given analytic true field, and the dashed lines depict the analysis. The upper part of the diagram shows the geopotential, and the lower part shows the wind component normal to the diagram. See text for further explanation.
Figure 2. As in figure 1, except that the length of the true correction field has been reduced to 10 degrees latitude.
1. Introduction

Linearized versions of the primitive equations permit wave solutions of two types, Rossby modes and gravity modes, the amplitudes of which are determined by the initial conditions. Large-scale Rossby modes resemble in structure and phase speed the transient wave motions observed in the atmosphere. Part of the large-scale gravity modes, however, have frequencies exceeding those of the observed large-scale flow by up to an order of magnitude. Thus, such large-scale high frequency gravity waves are not present in the atmosphere or at least they have exceedingly small amplitudes. On the other hand in numerical models of the large-scale flow based on the nonlinear equations large amplitude high frequency oscillations are set up, unless the initial data satisfy a delicate balance between the wind and mass fields. Such spurious frequency oscillations, obviously of the gravity wave type, have been called meteorological noise (Hinkelmann (1951)). (As an example see for instance Andersen (1977), Fig. 21). In models with built-in time and/or space dissipation the spurious oscillations die out after a certain period of time, the length of which depends upon how unbalanced the initial data was and upon the strength of the dissipation.

Even with present days' advanced observation and analysis systems large amplitude noise occur in the model integrations if the analysed fields are used directly as initial data. The primary reason for this noise seems to be errors in the analysed fields, but also model defects may contribute. (Thus, for instance truncation errors in the pressure gradient term over steep topography in θ-coordinate models...
may set up spurious gravity waves (Machenhauer and Hansen (1980))). The process of adjusting the analyzed fields with the purpose of obtaining initial fields that gives smooth integrations without the spurious high frequency oscillations is called initialization. Various initialization procedures developed up to the mid 1970's are reviewed, for example by Bengtsson (1975). As explained by Bengtsson none of these procedures are completely satisfactory.

In recent years a new approach called 'nonlinear normal mode initialization' has been developed for global primitive equation models. This approach, which is still under development, is now used at a number of forecasting centers and research institutions. The method offers a promising solution to one of the long standing problems in numerical weather prediction. A comprehensive review of the nonlinear normal mode initialization method has recently been published by Daley (1981). In the present paper we shall focus mainly on aspects not covered by Daley's review, especially the more fundamental concepts of nonlinear normal mode initialization. Also recent developments will be reviewed.

In section 2 the background for the classical initialization methods preceding the nonlinear normal mode method are considered briefly. Section 3 introduces normal mode initialization in general. In sections 4-6 the nonlinear normal mode method is considered in more detail, and in section 7 results of applications are summarized. Section 8 contains concluding remarks. In the presentation we shall assume the basic concepts as presented for instance in Daley's review to be known.
2. Classical initialization.

The classical static initialization procedures preceding the nonlinear normal mode method were based upon analysis of extremely simplified linearized versions of the primitive equations and upon midlatitude synoptic scale analysis.

Hinkelmann (1951) found that if the initial fields satisfied the geostrophic balance condition the amplitudes of the gravity waves would be small in a linearized shallow water model. The model considered was a mid-latitude, f-plane model. The basic state was stationary, with a mean height of the free surface equal to 10 km and a zonal flow in geostrophic balance. The perturbations were assumed independent of the north-south length coordinate. Guided by this analysis he suggested that initial fields in geostrophic balance should be used in PE-models in order to avoid excessive noise.

Charney (1955) tested the geostrophic initialization in a numerical integration of a nonlinear shallow water, mid-latitude, f-plane model with an equivalent mean height of 1 km. He found that initial fields in geostrophic balance gave rise to large amplitude gravity oscillations, whereas initial fields satisfying the nonlinear balance equations did not. Charney explained these results by the fact that although in both cases the initial divergence was zero its first derivative would be zero only if the balance equation was satisfied. Now, it is characteristic of gravitational motion that the horizontal divergence is not relatively small. He therefore expected that initial fields satisfying the balance equation automatically would exclude such motions initially, and also prevent them from being developed at later times.

Later on Phillips (1960) found by reanalysis of the linear model used by Hinkelmann (loc. cit.) that even in this model some gravity noise would be present with geostrophic initial conditions. The noise was found to increase with decreasing equivalent depth so that for higher internal modes in baroclinic models geostrophic initialization could be expected not to be satisfactory. His analysis showed that the amplitude of the gravity noise could be reduced substantially by adding a divergent wind
field to the nondivergent geostrophic field. This divergent wind field should be
determined from the quasi-geostrophic $\omega$-equation of the linear model. Like Charney
he generalized these results suggesting that for nonlinear models the initial rota-
tional and divergent wind fields should satisfy the nonlinear balance equation and
the quasi-geostrophic $\omega$-equation, respectively. He also pointed out, that in Charney's
test with the balance equation satisfied initially the quasi-geostrophic $\omega$-equation
was in fact also satisfied automatically due to the choice of a particular simple
initial streamfunction field. It should be noted, that both diagnostic equations used
in classical initialization are derived from the primitive equations by a scale analy-
sis valid approximately for mid-latitude synoptic scale systems. Within this approxi-
mation the equations imply that the local time derivative of divergence and of the
ageostrophic vorticity is zero. These properties imply that under certain special con-
ditions the classical initialization becomes approximately equivalent to first order
nonlinear normal mode initialization. The equivalence has been established only for
mid-latitude models and only when using normal modes defined by a linearized f-plane
model with a basic state at rest (Leith (1980), Daley (1981), Kasahara (1982)). The
approximate equivalence is a consequence of the fact that the f-plane Rossby modes are
nondivergent and geostrophically balanced. Subject to the same restrictions Hinkel-
manns geostrophic initialization is equivalent to linear normal mode initialization.

The classical initialization procedures are generally not satisfactory when
applied to present day's global or hemispheric multi-level models. The level of noise
is reduced by these procedures but it is still too high especially in the tropics, in
mountain areas and on the planetary scales. In operational practice usually only the
balance equation and not the $\omega$-equation has been used.
3. Normal mode initialization.

Normal mode initialization (NMI) methods may be seen as an extension of the classical initialization methods. In the latter methods one tried to generalize the constraints necessary to eliminate high-frequency oscillations in simplified linearized models. In NMI on the other hand one is working directly with a linearized version of the complete set of equations for the numerical model in question. The wave solutions to such a system are called the normal modes. Dickinson (1966) and Dickinson and Williamson (1972) developed procedures for determining normal modes of discretized models. It was the idea of Dickinson and Williamson (1972) to use such modes directly in an initialization procedure. They suggested to expand the initial data in terms of the model normal modes and then to set to zero the expansion coefficients for the unwanted modes.

This procedure, called linear normal mode initialization (LNMI), was tested by Williamson (1976) for a global grid point shallow water model. Starting with real data obtained from a 500 mb analysis all gravity modes (and the computational Rossby modes) were set to zero initially. Although as expected the noise was eliminated completely in an integration of the linearized model equations some noise was still left in an integration of the nonlinear model. This noise was found to increase with decreasing mean depth of the model atmosphere. Obviously this remaining noise was due to the terms in the model equations neglected in the linearized model.

One possible explanation of the remaining noise would be that the basic state chosen by Williamson, i.e. a basic state at rest, was too crude and that the linear normal mode initialization might be substantially improved if a more representative basic state was used.
The computational work connected with the determination of the normal modes is increasing with the complexity of the basic state. In order to keep the computations within practical limits it is necessary to be able to separate the zonal and meridional dependence of the normal modes. Thus the most complicated basic state to be used in a shallow water model is a stationary state varying only in the north-south direction. Machenhauer carried out some experiments with LNMI of a shallow water model. He found that even with a stationary basic state with a zonal flow equal to the zonal mean of the initial nondivergent velocity field almost the same results were achieved with the LNMI procedure as when a basic state at rest was used. Thus even with this more complicated basic state it seemed necessary somehow to take into account the nonlinear terms neglected in LNMI. Initialization procedures in which this is done are called nonlinear normal mode initialization (NNMI) procedures. Such procedures were developed independently by Baer (1976 and 1977) and by Machenhauer (1976 and 1977).

Machenhauer's NNMI procedure, which will be considered in more detail in the following section was based upon observations of the behaviour of the gravity modes during integrations of a shallow water model. These observations indicated that during a smooth forecast without high-frequency oscillations the linear and nonlinear part of the time derivative of these modes would balance each other closely. Consequently a smooth forecast could be expected if the gravity modes were determined in such a way that this balance was satisfied initially. To determine such initial data he used an iterative scheme by which the gravity mode expansion coefficients were successively changed while those of the Rossby modes were kept fixed. The iteration scheme was found to converge very fastly and when using the initial data obtained after only two iterations a time integration gave a smooth forecast with no indication of spurious high-frequency noise. The mean height of the shallow water model used in these experiments was 8.8 km corresponding roughly to the equivalent height of the external mode.
of baroclinic models and the normal modes were defined by a basic state at rest. In such a model all the gravity modes may be characterized as high frequency modes. Machenhauer expected that his method could be extended to baroclinic models and that it could be used successfully to eliminate the high frequency oscillations of such models, i.e. the high-frequency gravity modes of the external and first few internal modes. Such extensions were made by Andersen (1977), Daley (1979) and Temperton and Williamson (1979). We shall return to these initial tests of the method and to more recent tests of its performance in section 7.

Baer's NNMI method was described and tested originally for a simple $\beta$-plane model in which the latitudinal dependence of the variables was neglected. The method is building upon nondimensional model equations. Scaling parameters were chosen corresponding to mid-latitude synoptic scale motion and a mean depth of the model atmosphere of about 8 km. By assuming these scaling parameters characteristic of the model solution and especially by assuming the characteristic time scale to be the advective one, constraints on the initial data are obtained. Subject to the assumptions made an infinite succession of more and more accurate constraints necessary to avoid high-frequency gravity oscillations are obtained. Baer assumed the initial vorticity field to be known exactly. Given this field the constraints then determine the two other initial fields of dependent variables. A general description of Baer's method, without reference to any specific model was given by Baer and Tribbia (1977). In this description the normal modes of a model are divided into two groups: the high-frequency gravity modes of the model, called the fast modes and the remaining modes, called the slow modes. The Baer-Tribbia (B-T) initialization procedure is completely equivalent to Baer's procedure except that the initial slow modes are assumed to be known exactly as in LNMI and in Machenhauer's NNMI. Given the slow modes the B-T constraints determine the fast modes. The B-T zero order approximation is equivalent to LNMI. That is,
to setting the amplitudes of the fast modes to zero. The first order approximation is equivalent to one iteration of the Machenhauer iteration scheme starting with the zero order approximation. With the assumptions made this approximation should give a small value of the first derivative of the fast mode amplitudes. The second order approximation should give small second order time derivatives and so on.

Only very few results of tests of the B-T procedure have been published. We shall return to a more detailed description of the procedure in section 6.


The prediction equation for a single normal mode component may be written

\[ \frac{\partial}{\partial t} y_\gamma = i \sigma_\gamma y_\gamma + r_\gamma (yy), \]

where \( y_\gamma \) is the normal mode expansion coefficient of the dependent variables, \( \sigma_\gamma \) is the frequency of the normal mode component and \( r_\gamma \) is the expansion coefficient of the terms in the model equations not included in the linear system defining the model normal modes. \( Y \) is a vector including all the \( y_\gamma \)'s as elements. As indicated \( r_\gamma \) includes second order interactions between the normal mode components. We shall refer to the first and second term on the right hand side of (1) as the linear and non-linear contribution to the tendency, respectively.

In integrations of his shallow water model Machenhauer (loc.sit.) plotted separately time series of the two contributions to the tendency of the gravity modes. The normal modes used were determined from the linearized model equations using a basic state at rest. In an uninitialized integration he found that the two contributions were oscillating around a slowly varying value, which in many cases was found to be almost constant.

The oscillations of the linear part typically were large amplitude very nearly harmonic oscillations with a frequency very close to \( \sigma_\gamma \). Those of the nonlinear part on the other hand were typically relatively small amplitude oscillations of a much higher frequency. In the slowly evolving state around which the uninitialized inte-
gration was oscillating the linear and the nonlinear contributions seemed very nearly to balance to zero. Thus, for this slowly evolving state, for which we shall denote the expansion coefficient by \( \tilde{y} \), the gravity modes \( (\gamma \in \mathcal{G}) \) seemed very nearly to satisfy the balance condition

\[
\frac{d}{dt} \tilde{y}_{\gamma} = i \sigma_{\gamma} \tilde{y}_{\gamma} + r_{\gamma}(\tilde{Y} \tilde{Y}) = 0, \quad \text{for } \gamma \in \mathcal{G},
\]

where \( \tilde{Y} \) is a vector including all \( \tilde{y} \) as elements.

Since in the uninitialized integration the fluctuations of \( r_{\gamma}(\tilde{Y} \tilde{Y}) \) were observed to be small compared to those of the linear term and \( r_{\gamma}(\tilde{Y} \tilde{Y}) \) was observed to vary relatively slowly a good approximation should be obtained by substituting \( r_{\gamma}(0) \), the initial value of \( r_{\gamma}(\tilde{Y} \tilde{Y}) \), for \( r_{\gamma}(\tilde{Y} \tilde{Y}) \) in (1). Doing so we get

\[
\frac{d}{dt} y_{\gamma} = i \sigma_{\gamma} y_{\gamma} + \tilde{r}_{\gamma}(0),
\]

the solution of which is

\[
y_{\gamma}(t) = - \frac{\tilde{r}_{\gamma}(0)}{i \sigma_{\gamma}} + \left( y_{\gamma}(0) + \frac{\tilde{r}_{\gamma}(0)}{i \sigma_{\gamma}} \right) e^{i \sigma_{\gamma} t}.
\]

To the extent that this is a valid approximation for the gravity modes it explains the harmonic oscillations observed in the uninitialized integrations. It also shows that it is not sufficient to set \( y_{\gamma}(0) \) equal to zero in order to eliminate such oscillations, since many components were observed to have values of \( \tilde{r}_{\gamma}(0) \) significantly different from zero. Consistent with (2) we find that the amplitude of the harmonic oscillations in the solution (4) is set equal to zero by choosing

\[
y_{\gamma}(0) = - \frac{\tilde{r}_{\gamma}(0)}{i \sigma_{\gamma}}.
\]

In order to determine such initial values of the gravity modes values of \( \tilde{r}_{\gamma}(0) \) must be estimated. Machenhauer used the following iterative scheme

\[
y_{\gamma}(k+1)(0) = - \frac{r_{\gamma}^{(k)}(0)}{i \sigma_{\gamma}},
\]

where \( k \) indicates the iteration number and \( r_{\gamma}^{(k)}(0) = r_{\gamma}^{(k)}(\tilde{Y} \tilde{Y}) \) at \( t=0 \). As the deviations of \( r_{\gamma}(\tilde{Y} \tilde{Y}) \) at \( t=0 \) from \( \tilde{r}_{\gamma}(0) \) was observed to be small a good first guess \( r_{\gamma}(0) \) seemed to be \( r_{\gamma}(\tilde{Y} \tilde{Y}) \) determined from the initial unbalanced data. Only the gravity modes were changed in each iterative step. As mentioned before this scheme was found to converge very fast and initial data determined by only two iterations gave a smooth
integration without any signs of gravity oscillations.

The results from a 5 day integration also showed that even though the time derivative of the gravity mode coefficients were set (almost) to zero initially these coefficients did change during the integration, but in such a way that the balance between the linear and the nonlinear part remained very nearly satisfied.

For this shallow water model with a mean depth of 8.8 km the main difference between LNMI and NNMI was found for the ultra large scale gravity modes, that is the gravity modes with relatively low frequencies.

5. Extension to higher order balance conditions

Leith (1980) introduced the slow manifold, defined as the manifold of all model states that evolve slowly during a time integration. The manifold of states satisfying the Machenhauer balance defined by (2) was taken as a first approximation to the slow manifold. Machenhauer's experiments indicated that at least for the model considered it is a very good approximation. Leith argue that clearly (2) is not exactly consistent with the definition of the slow manifold, because slow dynamical changes will generate small nonzero values of $\frac{d Y}{dt}$. As one way to correct for this latter effect he introduces the second-order balance condition

$$\frac{d^2 \tilde{Y}}{dt^2} = (i \sigma_y) \tilde{Y} + i \sigma_y r_s(Y Y) + \frac{d}{dt} r_s(Y Y) = 0, \text{ for } Y \in F,$$

where $F$ represents the group of fast modes. Dividing (7) by $i \sigma_y$ we get

$$i \sigma_y Y + r_s(Y Y) + \frac{1}{i \sigma_y} \frac{d}{dt} r_s(Y Y) = 0.$$

Compared to (2) this balance condition includes an extra "correction" term, which becomes increasingly important the larger the time rate of change of $r_s(Y Y)$ is compared to the frequency $\sigma_y$.

Shortly after Leith's introduction of the second order balance condition Lorenz (1981) introduced a superbalance condition equivalent to

$$\frac{d^n \tilde{Y}}{dt^n} = 0, \text{ for } Y \in F,$$

where $n$ should be a large number. He applies it to a mid-latitude $f$-plane shallow
water model with forcing, diffusion and a bottom topography simulating the $\beta$-effect and a mean depth of 8 km. A low order system with only 9 time dependent variables is considered. To determine a balanced state satisfying (9) he is using a two-loop iterative scheme. Starting with a linear balance $y_\gamma^{(0)} = 0$ he determines at first $y_\gamma^{(1)}$ so that $dy_\gamma^{(1)}/dt = 0$. Using the final value as the first guess he then determine $y_\gamma^{(2)}$ so that $d^2y_\gamma^{(2)}/dt^2 = 0$, etc. until a set of $y_\gamma^{(n)}$ has been determined so that $d^ny_\gamma^{(n)}/dt^n = 0$. The procedure was found to converge rapidly and after $n = 3$ the computed values of $y_\gamma^{(n)}$ did not vary appreciably. When the superbalanced initial data was used the balance was found to be maintained during a numerical integration.

The extent to which the balance conditions of different order are satisfied during a model integration has been investigated by Errico (1982). He is using a highly simplified so-called sparse-spectral model. It is a mid-latitude f-plane two layer model in p-coordinates with $\omega = d\rho/dt = 0$ for $\rho = 0$ and $\rho = 1000$ mb. Thus only internal gravity waves corresponding to an equivalent depth of about 500 m are included. The model includes only 80 dependent variables each representing a band of spectral components. The model has parametrized dissipation and a constant forcing at a certain wavenumber band. In a 325 days integration he considers time mean values over the last 250 days of integration. Including all the gravity modes he computes the percentages of energy in these modes described by the solution to the balance equations

$$
\frac{d^n y_\gamma}{dt^n} = 0, \quad \forall \in G,
$$

for $n = 1, 2, 3$ and 4. The solution obtained for $n = 1$ is equivalent to the solution obtained by Machenhauer's scheme. Errico also computes the percentage of energy described by the so-called quasi-geostrophic solution. That is, the solution to (10) with $n = 1$ obtained by setting $y_\gamma = 0$ for $\gamma \in G$ in the nonlinear term of this equation. This solution is equivalent to the first order B-T solution. For atmospheric-like parameters he finds that an increasing amount of energy with increasing order $n$ is described by the solutions to the balance equation (10). The B-T first order
solution describes 80% of the energy, the Machenhauer balance solution 88%, the second order balance solution 96% and the fourth order balance solution 98%.

In order to see why the high-frequency modes in a solution belonging to the slow manifold should satisfy the super balance condition we shall consider the general solution to (1). Assuming $r_y(t)$ to be infinitely differentiable this solution may be written

$$ y(t) = (y_0(t) + \sum_{n=1}^{\infty} \frac{1}{(i\sigma)^n} \frac{d^{n-1}r_y}{dt^{n-1}} \bigg|_{t=0}) e^{i\sigma t} - \sum_{n=1}^{\infty} \frac{1}{(i\sigma)^n} \frac{d^{n-1}r_y}{dt^{n-1}}. $$

This solution is of course implicit, as the right hand side depends upon the expansion coefficients $y(t)$ of all the normal modes of the model. As such it says nothing about the actual time-evolution of $y$. If, however we observe harmonic oscillations of a coefficient $y_y$ with the frequency $\sigma_y$ then (11) strongly indicates how such oscillations might be eliminated. The condition for zero amplitude of the harmonic oscillation in (11) is

$$ y_0(t) + \sum_{n=1}^{\infty} \frac{1}{(i\sigma)^n} \frac{d^{n-1}r_y}{dt^{n-1}} \bigg|_{t=0} = 0, $$

which in fact is the limit of

$$ \frac{1}{(i\sigma)^p} \frac{d^p r_y}{dt^p} \bigg|_{t=0} = 0, $$

for $p \to \infty$, or the super balance condition. When (12) is satisfied initially it follows from (11) that it will be satisfied at any time. Note that if we assume the slow manifold solution to satisfy the superbalance condition (12) then (11) indicates that when using initial conditions satisfying a low order balance equation of order $p$ we should have oscillations with amplitude

$$ \sum_{n=1}^{\infty} \frac{1}{(i\sigma)^n} \frac{d^{n-1}r_y}{dt^{n-1}} \bigg|_{t=0} $$

Using the same model and data as in Machenhauer's experiments Rasmussen (1981) has shown that some normal mode components do in fact show harmonic oscillations of extremely small amplitude, when initialized using the Machenhauer scheme. He also showed that these oscillations could be eliminated using a scheme building upon the second order balance equation (8). Thus for this model (8) is a slightly better approximation to the superbalance condition than (2), but the only effect of using
initial fields satisfying (2) (approximately) is the small amplitude oscillations of
some gravity modes around the slow manifold solution.

6. The Baer-Tribbia scheme

Rather than repeating the derivation given in Baer and Tribbia (1977) (in the
following called BT) we shall use an alternative one leading to the BT scheme. In
BT the model equations are non-dimensionalized before transformed to normal mode
form. Here we shall non-dimensionalize the normal mode form directly. We write (1) as

\[
\frac{dY}{dt} = i\sigma_s Y_s + r_s(Y_s Y_s) + r_f(Y_s Y_f) + r_f(Y_f Y_f).
\]

Here the nonlinear term \(r(YY)\) is supposed to be quadratic. It has been split up in
three terms:

- \(r_s(Y_s Y_s)\) representing interactions between slow modes only,
- \(r_s(Y_s Y_f)\) " " slow and fast modes and
- \(r_f(Y_f Y_f)\) " " fast modes only.

We introduce the following scalings

\[
y_s = S y_s', \quad y_f = F y_f', \quad \sigma_s = S \sigma_s', \quad \sigma_f = S \sigma_f', \quad t = \nu_A^{-1} t',
\]

where index \(f\) and \(s\) refer to an arbitrary fast and slow mode respectively, and all
primed quantities are assumed \(O(1). \nu_A\) which is used to scale the time \(t\) is an
advective frequency supposed to be characteristic of slow manifold solutions. We also
need a scaling parameter for the interaction coefficients in the non-linear terms.

This scaling parameter is in BT assumed independent of the types of modes that inter-
act and to be the same for slow modes as for fast modes. Thus we introduce

\[
r_s = K_r s' \quad \text{for } \gamma = s \text{ or } \gamma = f.
\]

Insertion of these scalings in (13) for a slow mode component (\(\gamma = s\)) lead after
division by \(\nu_A^2\) to

\[
\frac{dY_s'}{dt} - i \frac{\nu_s}{\nu_A} \sigma_s' Y_s' - \frac{K_s}{\nu_A} \left( F Y_s' + F Y_f' + \frac{F}{S} Y_s' (Y_s' Y_f') + \frac{F}{S} Y_f' (Y_f' Y_s') \right) = 0.
\]
If we assume $\nu_A/\nu_A << 1$ it follows that the second term is small and that the sum of the last three terms must be $O(1)$. In order to make use of this conclusion to obtain an expression for $K$ in terms of the remaining scaling parameters we must have that one of the three terms is dominating the sum. The relative magnitudes of the three terms depend upon the ratio $F/S$. If we assume $F/S$ small, then the first of the three terms is dominating. If on the other hand we assume this ratio to be large then the last term dominates. We choose as a provisional assumption the first alternative in which case

$$K = \nu_A S^{-1}. \tag{17}$$

This is in fact the result we would have arrived at without any assumptions if $r_g$ represented a pure linear advection with the characteristic advection speed. Insertion of the scalings (14) and (15) with the value of $K$ given by (17) in (13) for a fast component ($\gamma = f$) gives after division by $v_F$

$$\epsilon \frac{d\gamma'}{dt} - i\sigma_f y_f' = -\epsilon \frac{S}{F} \left(n_f'(\gamma'_f,\gamma'_f') - \epsilon \frac{S}{F} n_f'(\gamma'_f,\gamma'_f) - \epsilon \left(\frac{F}{S}\right)^2 n_f'(\gamma'_f,\gamma'_f')\right) = 0, \tag{18}$$

where $\epsilon = \nu_A/v_F$. Assuming $\epsilon$ to be small and using the provisional assumption that $F/S$ is small it follows that the third term in (18) must be $O(1)$ as is the second term. Thus

$$F = \epsilon S \tag{19}$$

in agreement with our provisional assumption.

When (19) is substituted into (18) we finally get

$$\epsilon \frac{d\gamma'}{dt} = i\sigma_f y_f' + n_f'(\gamma'_f,\gamma'_f') + \epsilon n_f'(\gamma'_f,\gamma'_f') + \epsilon^2 (\gamma'_f,\gamma'_f'). \tag{20}$$

Subject to the assumptions made this is the scaled prediction equation for a fast mode component which must be satisfied if the solution belongs to the slow manifold. Only one nondimensional parameter $\epsilon = \nu_A/v_p$ appears in (20). For a given model this parameter may be related to a properly defined Rossby number. For the mid-latitude
model considered by Baer (1977) for instance this is the case if \( v_F \) is assumed to be equal to the inertial frequency.

Assuming that the slow mode expansion coefficients are given exactly we want to determine the fast mode coefficients satisfying (20) at the initial time. We write the solution as

\[
\mathcal{M}_f' = \sum_{j=0}^{\infty} \mathcal{M}_f'(j+1) \varepsilon^j,
\]

where the coefficients \( \mathcal{M}_f'(j) \) for \( j \geq 1 \) are to be determined. We substitute (21) in (20) and determine the first order approximation \( \mathcal{M}_f' = \mathcal{M}_f'(1) \) by neglecting all terms with a factor \( \varepsilon^j \) for \( j \geq 1 \) in the resulting equation. The second order approximation is determined by the first order approximation and by the equation obtained when neglecting all terms with \( \varepsilon^j \) for \( j \geq 2 \) as a factor. Continuing this process expressions for \( \mathcal{M}_f'(j) \) for \( 1 \leq j \leq k \) are determined for any number \( k \). Defining \( \mathcal{M}_f'(0) = 0 \) the \( k \)'th order approximation for the unscaled expansion coefficients is

\[
\mathcal{M}_f'(k) = S \sum_{j=0}^{k} \mathcal{M}_f'(j) \varepsilon^j,
\]

which is obtained by multiplying (21) by \( \varepsilon S \) and by adding the zero order approximation. This latter approximation is obtained from \( \mathcal{M}_f' = \mathcal{F}_y = \varepsilon S \mathcal{M}_f' \) by setting \( \varepsilon \) equal to zero. Substituting the expressions obtained for \( \mathcal{M}_f'(j) \) in (22) we finally get for the BT scheme

\[
\mathcal{M}_f'(k) = - \sum_{n=0}^{k} \frac{1}{(i\sigma)^n} \frac{d^{n-1}}{d\tau^{n-1}} \left( r_F'(\gamma^{(k-n)} \gamma^{(k-n)}) \right),
\]

by which the initial fast mode expansion coefficients are determined to any order of accuracy from given slow mode components. Actually in (23) for \( k \geq 2 \) small approximations has been made in order to get an expression including the whole nonlinear term \( r_F'(YY) \) and its time derivatives. The first few approximations determined by (23) are
The zero order approximation is equivalent to LNMI. The first order approximation give initial fields satisfying approximately Machenhauer's first order balance condition (2). With the third order approximation the second order balance condition (8) is approximately satisfied, and so on. It is seen that the BT scheme (23) with higher and higher value of $k$ gives successively more and more accurate solutions to the super balance condition (12). Subject to the assumptions made in the derivation of the scheme it gives the optimal way to determine the solution to the super balance equation. Note that if instead of the initial LNMI step ($y_f^{(0)} = 0$) the zero order approximation is taken to be the fast mode coefficients obtained from the analysis ($y_f^{(0)} = y_f(0)$) and if all time derivatives are neglected then the BT scheme reduces to Machenhauers scheme (6). Compared to the Machenhauuer scheme the practical difficulty with the BT scheme is the evaluation of the time derivatives. In practice they are evaluated numerically (Tribbia (1979)). So far only results of tests of the scheme with shallow water models and up to the second order approximation have been published (Baer (1977), Tribbia (1979), Ballish (1979) and Tribbia (1982)). For these models the scheme seems to work quite well. The assumptions made in the derivation may very well be satisfied in such models, but it is a question whether they are satisfied in more realistic global multi-level models. The results obtained with the Machenhauer scheme for such models do indicate that the initial LNMI step in the BT scheme is superfluous or maybe even harmful and that the terms involving time derivatives are in practice negligible. There is no reason to believe that the BT scheme is optimal unless the assumptions made in this derivation are
satisfied. It was assumed that \( \varepsilon \) is a small number. Thus, the group of fast modes is assumed to include only the very high frequency gravity modes. The assumptions made concerning the characteristic magnitude of the interaction coefficients are indeed very crude and it is a question whether they are valid in a complex baroclinic model. When the magnitude of the nonlinear term was estimated, contributions from the non-adiabatic terms in the equation were neglected. In particular release of latent heat is a very important factor especially in the tropics. Since latent heating is highly correlated with divergence fields such diabatic tendencies must be projected to a large extent on the gravity modes. This has been verified recently by Puri and Bourke (1982). It is obvious that the neglection of diabatic terms results in an underestimation of the magnitude of the gravity modes and therefore also of their importance in adiabatic nonlinear interactions. Also other factors of importance in a realistic baroclinic model are neglected in the derivation of the BT scheme. So it was assumed that only transient motions were present. The possible existence of quasi-stationary fields were neglected completely. It is well known that in reality as well as in the models we do have quasi-stationary fields which especially for the most large scale flow patterns are of significant magnitude compared to the transient fields. The normal mode coefficients are determined by a projection on the normal modes of the fields of deviation from the basic state chosen. In general these deviation fields will include a quasi-stationary and a transient part. The quasi-stationary part would be zero only if we were using a basic state equal to the model's quasi-stationary state and if this basic state was a stationary solution to the model equations. The normal modes used in practice are not defined by such a basic state, but by a basic state at rest with a given basic temperature profile, which is a function of the vertical coordinate only. It seems necessary to use such a simple basic state in order to achieve separation of variables and thereby keeping the computations within practical limits.
In the usual \(\sigma\)-coordinate models with mountains included this choice of basic state imply that even in a horizontally uniform standard atmosphere large amplitude deviation fields are obtained. With real data the deviation fields will include in addition also the quasi-stationary fields observed on isobaric surfaces. As a result the normal mode coefficients must contain a relatively large quasi-stationary component. The effect of taking into account such quasi-stationary fields in the BT scheme would be that the magnitude of the gravity modes were increased and that the weight of the time derivatives relative to the nonlinear term themselves would be decreased in which case the scheme would come closer to the Machenhauer scheme. Finally it should be mentioned that even the transient part of the ultra large scales are not in agreement with the assumptions as these have a characteristic frequency much larger than the advection one assumed in the analysis leading to the BT scheme.

7. The present state of NNMI

The extension of Machenhauer's NNMI scheme to baroclinic models followed shortly after its introduction. This extension and initial tests were carried out by Andersen (1977) and Daley (1979) for spectral models and for a grid point model by Temperton and Williamson (1979) (Published as Temperton and Williamson (1981) and Williamson and Temperton (1981)). The main results of these studies were that the iterative scheme was found to converge when applied to the gravity modes of the lower order vertical modes in adiabatic frictionless model versions. When including gravity modes of higher order vertical modes, that is modes with relatively low frequencies, the scheme was diverging. It was found that with a certain number of vertical modes included convergence was obtained in an adiabatic version, whereas the iterative scheme was diverging when contributions from physical parametrization terms were included in the nonlinear terms. The experiments showed however that in order to eliminate high frequency gravity mode oscillations from forecasts with a model including a full physical parametrization scheme it was quite adequate to use
data initialized in an adiabatic (frictionless) version of the model. It was also
found that reasonable vertical motion fields with good time continuity in connection
with extratropical depressions and realistic mountain induced vertical motion fields
were generated when using the adiabatic NNMI, all in agreement with synoptic experi-
ence and theory. As a consequence of these initial results adiabatic initialization
schemes were introduced at several centres. At ECMWF for instance the present
operational procedure is to use two iterations of adiabatic initialization including
all gravity modes of the five lowest order vertical modes of the 15 level model
(Temperton (1981)). The frequencies involved are indicated in figure 1 which shows
normalized frequencies (Ω is the angular velocity of the earth) as a function of
zonal wavenumber for equivalent depths 10 km and 100 m, approximately equal to those
of the first and fifth mode in the ECMWF model. (The dashed curve shows the advective
frequency corresponding to a solid rotation of 15º/day).

Fig 1.: Normalized eigenfrequencies of a shallow water model
for mean depths H = 10 km and H = 100 m.
The experiences gained with several years of application is that such adiabatic schemes work quite well at medium and high latitudes. Here the initialization scheme results in a complete elimination of high frequency oscillations in the forecasts, realistic and consistent initial divergence fields, and a high correlation between observed and model computed surface pressure tendencies (Bengtsson (1981)). Also in the tropics all high frequency noise is eliminated, but here undesirable effects of the adiabatic initialization are observed (Bengtsson (1981), Puri et.al. (1982)). The divergence fields are weakened substantially by the initialization, especially at the upper levels. For the large scale components at least these changes are clearly unrealistic as the Hadley and Walker circulations, especially their upper branches are almost eliminated. In the tropics also a much lower correlation between observed and model computed tendencies is obtained with the initialized fields compared to those found at medium and high latitudes. Also the spin-up time of the model is affected by the initialization. The spin-up time is the time it takes for the model to fully reach its own characteristic circulation regime. During the spin-up time in particular the strength of the divergence fields and total precipitation intensity is increasing. The spin-up time is found to be increased by the standard adiabatic initialization.

One reason for these undesirable effects is clearly the neglect of diabatic effects in the initialization procedure. The divergence and vertical motion fields obtained by an adiabatic initialization is only those forced by adiabatic nonlinear interactions. These fields are much too week in the tropics as here the vertical motions are forced significantly also by the diabatic heating, mainly that due to convection. Because of the too week initial divergence fields convection is suppressed at the start of the forecast, which explains why no spurious oscillations are observed and why the spin-up time is increased.

Recently significant progress has been made in experiments aiming at finding a satisfactory way to incorporate non-adiabatic effects in NNMI. In the initial experiments referred to above all gravity modes in a certain number of vertical modes were
initialized. This imply that when a sufficient number of vertical modes are included, so that all high frequency noise is eliminated, also the modes of low zonal wave numbers with relatively low frequencies are automatically included. It was demonstrated by Puri and Bourke (1982) that a moist convective adjustment scheme could be incorporated without any problems and with convergence of the iterative scheme if a certain low-frequency cutoff was used. In a nine-level spectral model four vertical modes were initialized, but for each zonal wave number only those with frequencies \( \geq \) the lowest frequency for the first internal vertical mode. The gravity modes in the second and third internal modes excluded by this low-frequency cutoff were found to be those which contributed to the main description of the Hadley circulation. Thus, when using the same frequency cutoff in an adiabatic initialization the Hadley circulation was retained. With the diabatic initialization, however, stronger vertical motions and in particular a stronger Hadley circulation was obtained. A further step was made by Kitade (1982) who managed to incorporate a complete physical parametrization scheme in NNMI. In an eleven-level model five vertical modes were initialized, but only those modes with periods smaller than 48 hours. Thus, also in this case the lowest frequency gravity waves were excluded. He is also using a modified iteration scheme obtained from the Machenhauer scheme (6) by reducing the corrections made in each step by a factor 2. With this underrelaxation scheme, in the following called the K-scheme, he obtains convergence also when non-adiabatic effects are included. He points out that the lack of problems with non-convergence of the iterative proces may be due not only to the use of the K-scheme but also to the particular vertical finite differencing scheme used in the model. With this non-adiabatic initialization only a slight reduction of the global RMS divergence was experienced compared to the analysis. The Hadley circulation was found to be well retained and in some tropical disturbances even an increase of the divergence was
observed. Forecast experiments showed that the non-adiabatic initialization efficiently remove high-frequency oscillations and that, compared to results with an adiabatic initialization, a more realistic and rapid development of some tropical disturbances was observed. It was also demonstrated that, compared to the adiabatic initialization, the non-adiabatic one reduces the spin-up time.

In the diabatic initialization experiments of Puri and Bourke (1982) as well as in those of Kitade (1982) the lowest frequency gravity modes in the retained vertical modes were excluded by a low-frequency cutoff. Wergen (1982) has shown how a non-adiabatic forcing of these large-scale, relatively low-frequency modes may be included in an initialization procedure. For the ECMWF 15-level model all gravity modes of the five lowest order vertical modes were initialized. For the gravity modes with frequencies lower than a certain cutoff frequency a non-adiabatic contribution was included in the non-linear tendencies. This contribution was determined as the time-average of the diabatic forcing during a preliminary 2-hour forecast from uninitialized data. With this non-adiabatic contribution kept constant in every iteration step the iteration was found to converge. The inclusion of the non-adiabatic effects resulted in a considerable reduction of the large-scale initialization changes in the tropics. So, the Hadley and Walker circulations were found to be well preserved. A positive impact on forecasts was also obtained. The spin-up time was reduced, a smaller error growth rate was observed in the tropics in the early stages, and after 5 days of integration some improvements were noticed also in mid-latitudes.

A possible reason for divergence of the Machenhauer scheme has been suggested by Phillips (1981) and Ballish (1981). It was observed at NMC that when too low-frequencies were included, divergence of the scheme occurred in connection with strong jet streams (personal correspondence). Apparently the divergence of the scheme was due to the strong advection. In order to explain these observations they consider a normal mode equation of the form
\[
\frac{dy_\gamma}{dt} = i\sigma_y y_\gamma - im_\omega y_\gamma + N
\]

where \( m \) is a positive wave number, \( \omega \) is a constant (positive) angular advection speed, and \( N \) is the remaining part of the tendency not included in the linear terms. As in practice a basic state at rest is used in NNMI a linear advection term as the second term on the right hand side will be included in the term \( r_\gamma \). Assuming \( N \) to be independent of the fast modes initialized and starting with a linear NMI step Machenhauer's iteration scheme is found to converge towards the solution

\[
y_\gamma = \frac{N}{i(\sigma_\gamma - m_\omega)} \quad \text{if} \quad |\sigma_\gamma| > m_\omega
\]

which is the correct solution to the balance condition \( dy_\gamma / dt = 0 \). If, however \( |\sigma_\gamma| \leq m_\omega \) the iterative scheme is diverging. This divergence occurs only for low-frequency modes with frequencies lower than the advection frequency defined by \( m_\omega \). It is easily shown that the K-scheme is stable for eastward propagating modes down to one third of the advection frequency. This downward extension of the range of convergence may, at least partially, be the explanation of the successful experiments of Kitade (loc.cit.)

Also Gollvik (1980, 1982) Thanning (1982a, 1982b) and Errico (1983) have considered the convergence of Machenhauer's iteration scheme for various simplified models. It is not possible here to go into details of their analyses. In general the analyses confirm as found in practice that a low-frequency cutoff is necessary, and that this cutoff frequency is increasing with the strength of the fields described by the slow modes. The mechanism in the divergence due to the inclusion of non-adiabatic effects is not clear. Thaning (1982b) using a low order version of a shallow water model includes a constant non-adiabatic forcing and finds that the cutoff frequency necessary for convergence is increasing with increasing strength of this forcing. For more details concerning the convergence problem the reader is referred to the above mentioned papers, two of which are presented at this Stanstead Seminar.
8. Concluding remarks.

The justification of NNMI is that spurious large-scale high frequency gravity oscillations are eliminated. The primary reason for such oscillations is believed to be errors in the analyzed fields. The initialized data should therefore be expected to give a more accurate representation of the true large-scale atmospheric state at the initial time. With the usual adiabatic NNMI the divergence fields obtained in extra-tropical regions are certainly more realistic after initialization, but those derived in the tropics, at least the most large-scale quasi-stationary fields, are evidently less realistic. As indicated by the experiments made by Puri and Bourke (1982) the undesirable effects of initialization for the quasi-stationary tropical divergence fields can be avoided by excluding the relatively low-frequency gravity modes contributing significantly to the description of these fields from the initialization procedure. The question is if initialization of these modes can at all be justified. It is not possible to see from the results presented by Puri and Bourke whether spurious oscillations are set up when leaving out these modes from the initialization. Experiments with the ECMWF adiabatic initialization scheme (personal correspondence) shows that this is the case when a low-frequency cutoff at a period of 12 hours is used. It should be noted however, that this cutoff is shorter than those used by Puri and Bourke for the ultra-long waves. The experiments of Puri & Bourke, Kitade and Wergen referred to in the preceding section do however indicate that it is possible to avoid the detrimental effects on the tropical divergence fields by including diabatic forcing in the initialization scheme. The results indicate that the tropical divergence fields obtained by an adiabatic initialization are more realistic than those obtained from the uninitialized analysis.

As mentioned previously the spurious gravity oscillations in uninitialized integrations is damped out in models including time and space dissipation. The length of the period of damping is probably in most cases less than 24 hours, at least for the higher frequencies. During this period the oscillations impair the
calculations of the mass and the divergence fields especially, and thereby the vertical motion and precipitation calculations. The usual adiabatic NNMI do improve the forecast results in this period, even in the tropics.

Errors due to spurious oscillations in a model integration become especially harmful when the model is used in a data-assimilation system, as for instance that developed at ECMWF (Lorenc (1981)). Here a short period forecast (at ECMWF 6 hours) is used to produce the first guess fields to a statistical ("optimum") interpolation scheme. One of the checks of the observations is to compare the observed values and to reject those which exceed the estimated forecast error by a specified amount. If the error is large due to large amplitude gravity oscillations (and if a realistic estimated error is used) this check becomes useless. Furthermore large estimated errors of the first guess fields reduces the weight of these fields in the final analysis and thereby the impact of information from earlier observations. A good initialization scheme is therefore a very important element in a data assimilation system, as long as the analysis system is not capable of producing sufficiently accurate and balanced fields. The present adiabatic NNMI scheme is definitely found to have a positive impact in assimilation systems. Due to the weakening of the tropical divergence fields, however, the analysis changes of these fields are almost exactly negated by the initialization changes in each cycle. Thus the net impact of the divergent wind field given by the previous observations is very minute. As a result especially the semi-permanent divergence fields at the upper levels are suppressed also in the analysis. The extension of NNMI to include also diabatic forcing should give a further substantial improvement in data assimilation systems.

As indicated already by the pioneering experiments with diabatic initialization this will probably lead to improvements in forecasts, not only in the initial stages and not exclusively in the tropics.

The present experience with NNMI building upon the Machenhauer balance condition indicates that all spurious high frequency oscillations can be practically removed
by this technique. The question is if further improvement can be obtained by initializing lower frequency modes. If spurious harmonic oscillations of such frequencies can be detected in forecasts then this would be a justification for such an extension. It seems likely however, that initialization of lower frequency modes would require the use of modes defined by more complicated basic states and also an extension to higher order schemes, including time derivatives of the nonlinear terms. If spurious regular oscillations of higher frequencies are not excited in forecasts then it seems unlikely that NNMI can be of any use of the initialization of such modes.

Although NNMI based on the Machenhauer balance condition seems to be able to eliminate the high frequency oscillations, such oscillations may exist in initialized forecasts without being detected. Perhaps, under special conditions and if so most likely in connection with rapidly developing systems. If this is the case then it is possible that an improvement may be obtained by using higher order schemes. Concerning the spin-up problem, which has been called the initialization problem of the second kind, NNMI seems not to be able to eliminate the spin-up time completely. One very important factor in this problem is the initial moisture fields, which are not affected by NNMI. In order to solve the spin-up problem improved analysis of moisture (and in some models also of liquid water content) is required. Such improvements must be expected to lead also to improved non-adiabatic NNMI results.

In the present paper we have considered only unconstrained initialization in which the slow modes are not changed, as these are assumed to be determined exactly, or rather as accurately as possible, by the analysis. The analysis systems are however designed in such a way that in a certain sense the analysis minimizes the error of the total fields. When the fast modes are changed the error of the total field is no longer a minimum. Different approaches has been suggested to correct for this effect of the initialization. In the variational NNMI procedure introduced by Daley (1978) the fields of analysis errors are assumed known and one tries then to
determine that slow manifold state which minimizes a weighted area mean-square deviation from the analyzed state. The weight functions to be used are supposed to be determined from the known fields of analysis errors. An extension of Daley's variational procedure is presented at this symposium by Tribbia (1983).

Other approaches which involve changes in the analysis and/or repeated analysis-initialization cycles are presented at this symposium by Phillips (1983) and Williamson and Daley (1983). Such correction procedures involve changes also in the slow mode coefficients. It must be expected that hereby significant changes, hopefully improvements, may be obtained in forecasts also beyond the initial stages as well as in the final initialized fields. It should be noted, however, that these procedures are building upon unconstrained initialization and that therefore their success depend upon the quality of the unconstrained initialization scheme chosen.

Acknowledgements

The author wishes to thank Mikael Rattenborg for reading, correcting and typing the manuscript, instead of the usual secretary, who broke her arm.

REFERENCES


Bengtsson, L., 1975: Four-dimensional assimilation of meteorological observations.

GARP Publications Series No. 15, Global Atmospheric Research Program, WMO-ICSU Joint Organizing Committee, Geneva, Switzerland.


Errico, R.M., 1983: Paper presented at the present Stanstead Seminar

Gollvik, S., 1982: Some aspects of non-linear normal mode initialization, using a
simple spectral shallow water model on the mid-latitude $\beta$-plane. Report DM-38,
Department of Meteorology, University of Stockholm.

Hinkelmann, K., 1951: Der mechanismus des meteorologischen lärmes, Tellus, 3,
285-296.

Kasahara, A., 1982: Non-linear normal mode initialization and the bounded deriva-

Kitade, T., 1982: Non-linear normal mode initialization with physics, Report
No 82-7, Florida State University, 57 pp.

Leith, C., 1980: Non-linear normal mode initialization and quasi-geostrophic

Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation

Lorenz, E.N., 1981: Attractor sets and quasi-geostrophic equilibrium,

Machenhauer, B., 1976: An initialization procedure based on the elimination of
gravity oscillations in a spectral, barotropic primitive equation model,
Ann. Meteor., 11, 135-138

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water
model, with application to normal mode initialization, Beitr. Phys. Atmos.,
50, 253-271

Machenhauer, B., and A.W. Hansen, 1980: Truncation errors in the pressure gradient
term in a $\sigma$-coordinate model and the effects of non-linear normal mode initialization.
GARP Int.Conf.on Prel. FGGE Data Anal. and Results, ICSU-WMO, Geneva,
349-357.

Phillips, N., 1960: On the problem of initial data for the primitive equations,
Tellus, 12, 121-126


Normal mode techniques have seen increasing applications in atmospheric and oceanic modeling recently, beginning some years ago with spectral representations of simple barotropic systems and leading to the initialization and possible analysis schemes to be discussed in this colloquium. The attractiveness of normal modes may be associated with the fact that often they represent functions which are as close as one can come to the basic characteristic solutions of the relevant prediction equations, albeit the linearized forms. Since normal modes do not (by current definition) exist for nonlinear systems, we must content ourselves with these linear solutions for the present. How appropriate these modes are to the atmosphere (and not just a model) is a tantalizing question.

Since the atmosphere is a highly linear system it clearly has no normal modes as defined relative to linear differential equations. Nevertheless, data and observations indicate that some well defined flow characteristics do indeed exist. Although difficult to isolate, Madden (1979) has identified a 16 day traveling wave in the large-scale planetary flow. Given the limitations in our current observational network, however, it is evident that not many of these characteristic "modes" of the atmosphere can be detected. There are nevertheless sufficient data to establish statistical characteristics of the atmosphere which may give some insight into its normal mode structure, if such exists. Unfortunately, statistical modes derived from atmospheric data are not associated with frequency, and are therefore difficult to relate to model modes. Yet, if the amplitudes of statistical modes persist in the atmosphere from
day to day, they may reflect important characteristic behavior. Moreover, if the structures of the statistical modes compare closely to model modes, some valuable interpretations and applications may evolve.

In pursuit of such possibilities, we have been searching for characteristic structures in atmospheric data on a planetary scale. This effort has recently been intensified with access to the carefully sampled and prepared FGGE data set. At present we are looking for both the characteristic vertical structures in the data as well as the latitudinal structures. In keeping with the current climate of spectral modeling, our procedure to determine statistical structures is as follows. Hemispheric data of winds are converted to stream field values (the rotational component) in each pressure surface and expanded for each reporting period into the coefficients of a truncated solid harmonic series. For each planetary wave, the vertical structures of all included Legendre polynomials may be developed into a matrix. This matrix may be expanded to include the individual matrices for each recording period. We develop a covariance matrix from this final matrix. The eigenvectors of this covariance matrix reflect the characteristic vertical structures of the specified planetary wave, and the corresponding eigenvalues reflect the relative importance of these vectors to the total structure.

The entire data set for the stream function, given in terms of expansion coefficients, is projected onto its vertical characteristic structures, planetary wave by planetary wave. Thus the data set for the entire time record converts to a set of expansion coefficients in each vertical mode (with the modes determined from the statistics of the observations) rather than as given originally at each pressure surface. For each planetary wave and each vertical mode (structure), a latitudinal profile may be developed from the Legendre polynomial expansion. Evaluated at discrete latitudinal points, these profiles may be expressed as a vector. With one such vector for each recording period, a matrix which includes all time records is generated and its covariance matrix established. The eigenvectors of this
covariance matrix represent the characteristic latitudinal structures for the given planetary wave and vertical mode, with the corresponding eigenvalues identifying the relative significance of each structure (eigenfunction). The intent in transferring to vertical modes is to bring the analysis as close as possible to shallow water theory; i.e., hopefully each vertical mode represents the structure of that part of the atmosphere with given equivalent depth.

The above analysis procedure was applied to the available temperature data as well as the stream field. Additionally, since we are unable to represent the characteristic structures in terms of frequency, we have in a separate analysis removed the time mean from the data, thereby describing the properties of the transient motion and temperature fields. Whether the structures derived from this transient data set are meaningful remains to be determined. The reason for our implied skepticism arises from the observation that during the time averaging period many waves or modes may arise and disappear, and the statistics remaining from this process may not be informative.

The data utilized in this study are taken from the FGGE SOP1. Since we had some interest in the shorter planetary scales, only the northern hemisphere data were considered; the global data will be analyzed at a later date. Two analyses of the level II data were utilized, the first analysis by NMC (IIIa) and the more recent and careful analysis be ECMWF (IIIb). Based on the gridding of these analyses, the spectral data were truncated triangularly at wave 31, but only the first 25 waves are discussed. In all cases, there are data at 12 pressure levels from 1000-50 mb. It should be noted that analyzed fields include information introduced by the analysis procedure, and could modify to some extent the "true" atmospheric information taken from the station data. Hopefully such distortions are small, since we cannot eliminate them from our data, but we must nevertheless be alert to such shortcomings in our final interpretations.

Characteristic vertical structures have been developed previously from atmospheric data. Holmstrom (1963) discussed
structures determined from station data whereas Bradley and Wiin-Nielson (1967) showed structure as a function of planetary wave. Structures associated with various models have been described by Kasahara and Puri (1981), Temperton and Williamson (1981) and Baer and Sheu (1982). Since most P.E. models have strong similarity, model modes also show commonality. Some vertical structures determined from the FGGE data analysis prepared by this study are described in Fig. 1. The first six vertical modes for both the IIIa (FIGGYA) and IIIb (FIGGYB) analyses for all planetary waves up to wave 20 are depicted. For mode 1, structures of waves shorter than wave number three all have similar form for both the IIIa and IIIb data, but the IIIb structures do not have a kink at 70 mb. This similarity of structure for the shorter waves is evident for all six vertical modes. There are, however, significant differences between the structures for the IIIa and IIIb analyses for the internal vertical modes. Note for example how the maximum and minimum in the structures of mode 4 are shifted lower in the atmosphere for the IIIb analysis. Fig. 1 represents structures derived from the total flow. Structures representing the transient flow do not show the strong similarity amongst the planetary waves for each vertical mode as seen on Fig. 1. Eigenvalue analysis indicates that most of the amplitude is in the first vertical mode (external), and in this mode the zonal flow dominates. For the remaining vertical modes, the eigenvalue amplitude gradually decreases with increasing mode number but is not strongly dependent on planetary scale.

Statistical profiles of latitudinal structure, taken from atmospheric data in various vertical modes, have been shown by Bradley and Wiin-Nielson (1967), but only for a few modes. Theoretical functions (Hough) which satisfy linear models have been described for the atmosphere by Kasahara (1976) and for a P.E. model by Baer and Sheu (1982). In our analysis of the FGGE data, we have developed these structure functions for the longest planetary waves ($m = 1, 2, 3$) and for the first three vertical modes. Examples of these structures may be seen on Fig. 2. The first six latitudinal structures for each planetary wave and
vertical mode are shown. Latitude increases along the ordinate. No specific significance is suggested by these structures, nor do they compare in any apparent way with either Hough modes or model modes. Since no frequency information can be readily associated with these structures, similarity to specific Hough modes would only stand out if that Hough mode were predominant in the statistics of the flow. Eigenvalues related to these vectors indicate that the first latitudinal structure dominates for all three planetary waves of the external (first) vertical mode. This is not the case for the internal vertical modes, where the first three latitudinal profiles play a significant role.

A number of applications for the statistical analyses discussed herein come to mind. Possibly one can use the characteristic vectors as expansion functions for model prediction. This was attempted some years ago for the barotropic vorticity equation by Karhila and Rinne (1974), and is currently under experiment with Hough harmonics by Kasahara at NCAR. The derived structures may give some insight into model fidelity by comparing model modes to observed structures. Additionally, by projecting model output onto the statistical structures, prediction skill may be more effectively diagnosed. Statistics in scale ranges outside of a model's resolution may be utilized to develop suitable parameterization. Since the statistical characteristic vectors derived in this study represent the most significant variance of the specified variable by relatively few terms, projection of observational data onto these vectors should provide substantial data compression and thereby minimize data storage. Indeed careful study of these vectors may lead to alternate and possibly more efficient and effective observational schemes for atmospheric parameters. Finally, the observed statistics discussed herein may be of some assistance to the initialization problem.
References


Legends

Fig. 1: The first six characteristic vertical structures of the stream field taken from both the FGGE IIIa and FGGE IIIb analyses and presented for each planetary wave.

Fig. 2: The first six characteristic latitudinal profiles of the stream field taken from the FGGE IIIb analysis for 3 vertical modes and 3 planetary waves.
Initialization Experiments with the NMC Spectral Model
Bradley A. Ballish
Development Division, NMC

In this extended abstract, Machenhauer (1977) and Baer-Tribbia (1977), henceforth referred to as BT, initialization schemes are first briefly compared theoretically. Various versions of the above schemes are discussed and compared by application to a forecast model. The model used is the NMC global spectral model, Sela (1980), with 12 vertical levels and a rhomboidal truncation to 24 waves. Earlier work on initialization with the model can be found in Ballish (1980).

The Machenhauer initialization scheme involves an iterative procedure that attempts to make the initial time derivative of high frequency gravity modes zero. Given a gravity mode tendency $\dot{Y}$, we set $\Delta Y = -\dot{Y}/i\omega$, where $Y$ represents a single gravity mode, $\Delta Y$ its change in one iteration, and $\omega$ its frequency. More than one iteration may be employed as nonlinearity prevents the above $\Delta Y$ from making the new value of $\dot{Y}$ exactly zero. The BT procedure gives approximately the same results as the Machenhauer procedure. The main difference is that the BT method first sets the amplitude of the high frequency modes to zero. Then the gravity modes are adjusted to be in balance with their nonlinear forcing so that there should be no high frequency activity during the ensuing forecast. To first order, the BT procedure is equivalent to gravitational zeroing followed by one iteration of the Machenhauer procedure. When carried beyond first order, the BT scheme deliberately makes the initial gravity mode tendencies small but nonzero. As weather systems evolve and move, the balanced flow has ageostrophic fields which change with time. It is this natural slow time behavior of the gravity modes that the BT procedure
attempts to give rather than having $\dot{Y}=0$ exactly.

Two modifications of the standard BT initialization are introduced. First, the gravitational zeroing step is modified. Instead of first zeroing out the amplitude of all high frequency modes, the gravity modes of the model with standard atmosphere data are retained. This is done because even in a standard atmosphere, there are gravity modes in mountain areas due to the use of sigma coordinates. This modified zeroing results in a better balance than pure zeroing, especially in mountain areas. However after the first order correction of the BT procedure, the orographic induced error from the pure zeroing is largely corrected, so this special modification need not be critical.

A second modification applies only to first order BT initialization. Without the cost and possible convergence problems associated with second order initialization, we can easily improve upon the first order results. It is found useful to perform one extra additional iteration of the Machenhauer scheme for the external mode only after one has finished doing the first order initialization. The changes produced by this extra iteration are small in amplitude but have a significant effect on reducing high frequency surface pressure noise. Because the external gravity modes are, for the most part, distinctly faster than all internal gravity modes, one can use scaling arguments to show that this extra iteration is worthwhile and valid even though it is not fully correct to second order accuracy.

We find that full second order BT initialization results in less noise than first order BT initialization with or without the additional external mode correction. It is found that the second order term involving the Rossby mode tendencies is a significant factor in the improvement
of the second order results. However, the second order results appear
to have little improvement over first order initialization in the accuracy
of the ensuing forecast. In addition, one must be more careful about how many
vertical modes are initialized to second order. In some cases, second order
initialization for 6 vertical modes has a slightly larger divergence tendency
at jet level than does the first order initialization, although the overall
balance is better with the second order scheme. This may be due to possible
nonconvergence of the full BT procedure (or the larger tendencies may be
correct).

In comparing Machenhauer and BT initialization results, we find that the
BT scheme is more capable of reducing high frequency noise and provides a better
mass-motion balance. If we perform strict first order BT initialization, the
mass-motion balance may be better than what is possible with the Machenhauer
scheme, but the surface pressure noise is usually larger. Adding on the extra
external mode correction to the first order BT results tends to give an overall
better result than the Machenhauer procedure. The biggest difference between
the usual Machenhauer initialization and the BT procedure is the gravitational
zeroing step of the BT method. Although it is possible that the zeroing may
destroy a fine gravitational balance provided by the analysis as well as
having some problems in mountain areas, the zeroing step appears to be
generally beneficial. The most significant aspect of the zeroing occurs
in jet stream areas. If the analysis produces a gravitational imbalance
which is acted on by strong advecting winds in a jet, the conventional
Machenhauer scheme has been found to be incapable of correcting the
imbalance. When the NMC spectral model was implemented, it used two
iterations of Machenhauer initialization for 6 vertical modes. This ran
operationally for some time without any apparent serious problems. Then
in the fall of 1980, the initialization showed some poor behavior in strong jet streams. Occasionally, the analysis would produce a somewhat nonmeteorological appearing imbalance between the mass and wind fields in jet areas. The initialization, in some cases, made the imbalance and appearance worse. This was due to nonconvergence in jet areas coupled with the imbalance introduced by the analysis. In such cases, the BT procedure tends to give more meteorological and balanced results, but the results may have some significant differences with the data and the analysis. Because of this possible jet stream problem with 6 vertical mode Machenhauer initialization, we now use Machenhauer initialization for only 4 vertical modes in NMC operations.

For forecast skill, the BT procedure has the advantage that a large number of vertical modes can be safely initialized. In a number of experiments, we have found that by using BT initialization for more vertical modes, we improve the model's balance and forecast skill. Usually the improvement in forecast skill is small, but in some cases it can be significant. In a number of experiments of forecasts for 6 to 24 hours, first order BT initialization for 12 vertical modes plus the external mode correction usually gave the best results. In one 24 hour forecast, the BT initialization for 12 vertical modes produced an approximate 6% reduction in RMS height and wind errors and the model's RMS divergence tendency was roughly cut in half at hour 24 of the forecast, compared to the operational initialization results. The above encouraging results with 12 vertical mode initialization occurred even though such initialization has problems with large boundary layer friction.
References


A Comparison of Two Static Initialization Methods in Data Assimilation

Edward H. Barker
Naval Environmental Prediction Research Facility
Monterey, CA 93940

1. Introduction

The primary purpose of initialization procedures is to control unusually large gravity noise that may adversely affect the assimilation of data. Unfortunately, initialization methods may introduce some problems of their own, such as the suppression of precipitation in the tropics. It seems reasonable, therefore, to expect that the vertical motion of the forecast first-guess may be a good method of obtaining vertical motion. Especially since it is consistent with the moisture patterns developed by the model during assimilation.

Leith (1981), using large-scale quasi-geostrophic theory, has shown that scaling of the nonlinear normal mode (NLNM) method is similar to that of the balance and omega equation combination. Consequently, if the forecast first-guess divergence is sufficiently accurate, the balance equation may be competitive with the more difficult NLNM methods.

The primary purpose of this comparison is to evaluate the possible benefits of the vertical motion of the forecast first-guess fields. Additionally, the manner in which the balance equation method measures up to the NLNM method is to be evaluated. A brief description of the data assimilation method, balancing method and the evaluation of several experiments are presented in the following.

2. The Data Assimilation System

A simplified version of the Navy Operational Global Atmospheric Prediction System was used in this study. The analysis was univariately performed using successive corrections in three dimensions. Analysis of wind and geopotential were produced for each of the standard levels on a 2.5 by 2.5-degree grid. Balancing was performed on pressure surfaces in the balance equation method and on model coordinates in the NLNM method. The forecast model, which was originally developed at UCLA as a general circulation model (Arakawa and Lamb, 1977), has six levels and a horizontal resolution of 4° longitude by 5° latitude. The top of the model was set at 50 mb. Each twelve hours, the model was updated using the FGGE level IIa data received during the first two weeks of November, 1979.
3. The Balance Equation Method

The balance equation method is best described using a vertical vector form,

\[ m(\psi, \phi) = \nabla \cdot (f \nabla \psi) + A - \nabla^2 \phi, \]  

where, for example, the vector \( \phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} \) contains horizontal fields of geopotential at each level. \( \psi \) is the stream function and \( A \) is the nonlinear component of the balance equation. The data and equations are then transformed into empirical orthogonal function coefficients and then the functional

\[ I = \int_R (\hat{\phi - \hat{\phi}})^2 + \beta (\hat{W - \hat{W}})^2 + 2\lambda \hat{m} \, dR \]  

is minimized using the Sasaji (1958) method. \( \beta \) is the weight on wind, \( \lambda \) is the Lagrange multiplier and \( R \) is the spherical domain over which the integral is minimized. The symbol \( \hat{\phi} \) is used to designate the transformed variables, i.e.,

\[ \hat{\phi} = E^{-1} \phi, \]  

where \( E \) contains the empirical orthogonal functions of the geopotential analysis computed using the method of Obukhov (1960).

The balanced fields are constructed using

\[ \phi = \sum_{i=1}^{N} \phi_i E_i \]  

and

\[ \nabla \times \mathbf{V} = \sum_{i=1}^{N} (\nabla \times \mathbf{V}_i) E_i \]
where \(N\) is the number of functions used to reconstruct the variable in physical space. \(N\) equal to four avoids some unrealistic temperatures that form in the lowest layers.

The divergence of the first-guess fields is retained in the balanced fields using two different methods as follows.

(a) The balance is performed on the analysis increments (the difference between the analysis and forecast first-guess), then the balanced increments are interpolated to model coordinates and added to the forecast first-guess.

(b) The balance is performed on the analysis and then the wind is converted to a rotational correction using

\[
(\nabla \times \mathbf{v})' = (\nabla \times \mathbf{v})_B - (\nabla \times \mathbf{v})_F \tag{7}
\]

before interpolation to model coordinates.

4. The Normal Mode Method

The normal mode methods were designed after Williamson and Dickenson (1976), Temperton and Williamson (1981), Williamson and Temperton (1981) and Machenhauer (1977) using the Tribbia (1982) variational method. The magnitude of the array inversion was too large to make the weights vary with both latitude and longitude. The two methods tested are as follows:

(a) The weight on the geopotential was 10 poleward of \(\pm 30^\circ\) latitude and 0.5 elsewhere.

(b) The weight was 1 everywhere.

The Machenhauer (1977) balance was applied to the external and first internal gravity modes. The equivalent depths of these modes are 7, 874 m and 784 m for a globally averaged temperature profile.

5. Results

The methods described above were tested in data assimilation experiments lasting five days. The effectiveness of the methods were evaluated from the way that they controlled gravity noise, generated realistic divergence, allowed precipitation during the early forecast hours, and produced accurate forecasts. The evaluation of each of these criteria is described in the following.

(a) Noise Control

The root-mean square (RMS) surface pressure tendencies were computed for the different methods described above. From these plots, it is clear that the nonlinear normal mode method was superior to the balance equation method in forecasts excluding the
heating parameterization. Surprisingly, the forecasts that included heating produced about the same RMS surface pressure tendencies for the two methods.

(b) Divergence and Precipitation

A visual comparison of the forecast first-guess divergence and that computed from the NLNM methods showed that the divergence fields were not significantly changed by the NLNM initialization. Precipitation during the first six hours of the forecast was found to be very sensitive to the divergence contained in the initial fields, where no divergence resulted in very small precipitation rates. All of the methods described above contained divergence in the initial fields, and the precipitation patterns were nearly identical. Furthermore, the patterns of the first six hours were consistent with those of the second six hours.

(c) Forecast Verification

When compared to observations, the forecasts from the balance equation method produced slightly better scores than the NLNM method. This may be partially due to the fact that the features from the balance equation method were smoother.

6. Summary

The gravity noise in forecasts initialized with the NLNM method is much smaller than those initialized with the balance equation, except where heating effects are included in the forecasts. The heating generates enough gravity noise that it masks nearly all of the benefits of using the NLNM method. This may be partially corrected by changing the way that heating is computed in the model.

Divergence from the forecast first-guess fields was sufficiently accurate to produce consistent precipitation forecasts, even during the early hours of the forecast.

The balance equation, which is considerably easier to apply, makes it easier to impose variations in the weights of the variational method. Consequently, the forecasts initialized with the balance equation verified against observations slightly better than did the NLNM method.

Much more testing is required before any conclusions can be made on the different initialization methods and the advantages of using the forecast first-guess divergence.
7. References


I. INTRODUCTION

There is no doubt that the Global Weather Experiment of FGGE will be regarded as a milestone in the development of meteorology. For the first time a truly global data set has been produced and the experiment has applied the most complete set of atmospheric data ever obtained. No doubt it will give the scientists an exceptional data base for research directed at improving methods for weather forecasting and understanding of climate, and thereby justifying the costs and efforts spent on the experiment.

We will here present some recent results obtained at ECMWF with the emphasis on observing system experiments. The processing of the FGGE II-b data at ECMWF has been done with the ECMWF operational forecasting system, Bengtsson et al. 1982a.

II. FGGE DATA

The amount of data during FGGE was substantially higher than what is received operationally on the GTS. Different comparisons undertaken during the SOPs indicated that in the average there were about 30% more rawinsondes, 3 times more aircraft data and about 3 times more satellite sounding data. However, as will be demonstrated, it is not only the amount of the data which is important. More important is their spatial distribution, in particular in areas where the atmospheric variability is high.

The quality of the data during FGGE has been carefully assessed. Observations have been compared by accurate reference observations, such as high quality radiosondes. Observations have also been evaluated by examining the fit to the analysis. Table 1 summarises the results of such comparisons. The estimated accuracies are representative of the synoptic scale and the error estimate includes therefore contributions from subsynoptic scales in addition to instrumental errors. It can be noted that the quality of the data from the satellite observing system is generally less good than the rawinsonde data and aircraft data. The satellite wind data in particular have large errors at higher levels due to difficulties of establishing the correct height of cloud pixels. However, it should also be noted that the satellite sounding data are more accurate in the stratosphere than are the radiosonde data.

III. OBSERVING SYSTEM EXPERIMENTS

A comprehensive presentation of the FGGE research at ECMWF can be found in Bengtsson et al. 1982b. We will here concentrate the presentation on observing system experiments. Certainly one of the most important tasks of the meteorological services is the
establishment of an operational observing network that can satisfy the accuracy requirements of large-scale numerical weather prediction in all areas of the globe. The observing systems during FGGE did, in many respects, satisfy these requirements, and it is therefore essential to evaluate their performance and establish to what extent they can be a part of a future operational system. Figure 1 shows the interpolation error as calculated from the ECMWF analysis system of the 500 mb height field in January 1982 for the Northern Hemisphere for a typical distribution of surface and operational upper air soundings.

The application of the ECMWF analysis system implies that the interpolation error is calculated from a three-dimensional multivariate scheme and hence interpolation takes place in the vertical dimension as well. The wind observation affects the height field and vice versa through the geostrophic relation which has been used in the derivation of the multivariate equation, Lorenc, 1981. As can be seen from the figure, there are two main areas where the interpolation area is large; one over the North Pacific and the other over the Central North Atlantic. The reason for the large interpolation errors in these two areas is a combined effect of an unsatisfactory data coverage and a higher variance. The low values in the tropics, which are as data sparse as the ocean areas in middle latitudes, are thoroughly due to the low variance of the 500 mb height field. Similar calculations for other levels and other parameters have been done.

It is to be expected that an improvement of the data coverage in data sparse areas will affect the predictive skill in general and immediately downstream in particular. The improvement is likely to be found in situations when there is a great meteorological activity in these areas. The Pacific region is especially sensitive since the data critical region is extended along the direction of the westerly flow and weather systems can, therefore, go undetected for some time before they can be properly analysed. Due to an extensive coverage of satellite and aircraft data as well as to more observations from the conventional WWW network, the 500 mb interpolation error during FGGE was substantially reduced, Figure 2. This was particularly the case in November 1979, when there were two orbiting satellites in operation. The interpolation error in Figure 2 refers to a six-hour time interval around 00Z 10 November 1979. The data coverage was representative for the month of November. As can be seen, there are no areas where the error is above 40 m, and the average 500 mb error has been reduced to 11 m. The corresponding reduction of the Southern Hemisphere extratropics is from 45 m to 21 m.

However, the kind of network assessment described above can only provide a first crude estimate. In order to obtain the full effect of the data, in particular the non-synoptic ones, it is necessary to evaluate the observing system by means of four-dimensional data assimilation and numerical prediction experiments. One such experiment will constitute the control case and contain all the reference data. In the other, the specific observations under investigation are excluded or added. Due to the fact that we are using prediction models for the time interpolation as well as for the adjustment between the different variables, the
observations must be assimilated for several days before the forecast can commence.

An international research programme has been started under the auspices of the JSC, to undertake impact experiments with the aim of evaluating the FGGE observing systems. As was demonstrated during the recently held JSC Study Conference, 1982, positive impact is shown from the FGGE observing system, although considerable problems do exist with the proper evaluation of the experiments. Furthermore, data errors of a systematic nature and deficiencies in the models and the data assimilation schemes generally reduce the impact.

An observing system experiment has recently been carried out at ECMWF. This consists of a set of seven prediction experiments, all using different data sets, Table 2. The reference experiment has been the case where all the FGGE data have been included. For each of the seven data sets, 8 days of the data assimilation have been done, starting 00Z 8 November 1979, all using climatology as a first guess. Ten-day predictions have been done starting from 00Z on 10, 11, 13 and 16 November 1979. Although the experiments have been set up to use FGGE data as a reference data set, the result of the experiments can also be analysed using any other data set as a reference data set. This can preferably be done by regarding system F, where we have removed all aircraft and satellite data, as another reference system, see further Table 2.

Table 3 summarises the result from the seven forecast experiments. The average predictive skill is expressed in days it takes for the anomaly correlation to reach a certain prescribed percentage, Bengtsson, 1981. We have selected 80 % as a figure to indicate a good forecast, 60 % to indicate a useful forecast and 40 % to indicate a possibly useful forecast.

Table 3a, which gives the results for the Northern Hemisphere, shows that aircraft data and satellite data are crucial for the forecast, and a considerable reduction in the skill can be seen when these observations are removed. The predictive skill for useful forecasts, for instance, is reduced from 6.8 days to 5.1 days. It is of particular interest to note that in this case the corresponding reduction in predictive skill is larger than the reduction which is found in system G, where all conventional upper air data have been removed from the FGGE system (6.8 days to 5.7 days). It can furthermore be seen that the different satellite systems and aircraft systems per se in this episode are almost as good as all the systems taken together.

The result for the Southern Hemisphere, Figure 3b, demonstrates strongly the important role of the satellite observing system. Both the good and the useful forecasts are improved by more than 1.5 days. There are hardly any aircraft observations at the Southern Hemisphere during this period, which explains the fact that no impact can be found. It should be stressed that none of these experiments evaluate the effect of buoys since the FGGE buoys are included in all of them. Further improvements would have been expected if the buoys had been excluded from system F.
We will next make a synoptic examination of four of these systems; system A, B, C and F, with the emphasis on the performance in the data sparse region in the North Pacific and the effects downstream from this area. Very active synoptic development took place during this time in this area and we may, therefore, expect sensitivity to changes in the observing systems. Figure 3 shows the 500 mb height field over the North Pacific for these four data sets for 00Z 11 November 1979. It can be seen that A, B and C fields are very similar, having an active trough in the quadrant 50°N to 40°N, 170°W to 160°W. This trough cannot be found in F, which has a more straight pattern in this area and where the active part of the trough is further to the East. A careful evaluation of the data available to the analysis schemes shows that the structure of the trough is well defined by the aircraft data, as well as by SATEM and SATOB data. In this particular case, therefore, the analysis will not change very much if not all of them would be present.

However, there are examples of other situations where this is not the case and where the existence of one of these observing systems is crucial to the forecast.

We will next examine the 48-hour forecast from these four different initial states. As can be seen from Figure 4, the trough moves eastward under intensification, and a well-defined cut-off low can be seen at 40°N, 150°W. The phase of this trough is reasonably well predicted in the forecasts A, B and C, although they all underestimate the amplitude. The forecast F, on the other hand, is substantially worse with an erroneous low at 30°N, 170°W. Major prediction errors can be seen from the error map in Figure 4.

Although we can in this case demonstrate an overall improvement of the predictive skill, as expressed by the anomaly correlation as well as by an improvement in the synoptic patterns downstream the data sparse areas, it must be stressed that these kinds of impact tests very much depend upon the model and the data assimilation system as well as on the particular weather type. Many more experiments with different models and data sets are needed before any firm conclusion can be drawn. At this moment we may only take comfort from the fact that we have noticed a clear positive impact from space and aircraft observing systems.

IV. POTENTIAL IMPROVEMENT DUE TO BETTER DATA ASSIMILATION SCHEMES

We would finally make some general comments about potential improvements in numerical weather prediction as a function of three factors; (i) accuracy of the initial state, (ii) model accuracy and (iii) data assimilation frequency. As has been demonstrated by Leith (1978, 1980), the growth of error in time for short range forecasts can approximately be expressed by the following linear equation.
\[ E = \alpha E + S \]  

which solution can be written

\[ E(t) = E_0 + (E_0 + \frac{S}{\alpha}) (e^{\alpha t} - 1) \]  

\( E(t) \) = error variance  
\( E_0 \) = initial error variance  
\( \alpha \) = growth value = 0.55 day\(^{-1}\)  
\( S \) = model error source rate

Equation 1 can also be used to estimate the reduction of the error which is obtained by successively assimilating initial states with an error variance \( E_0 \) into a numerical model.

If \( \tau \) represents the fixed pseudo-time interval of the data assimilation cycle, a numerical solution of equation (1) can be written

\[ E_{\tau,n} = E_n (1 + \alpha \tau) + S \tau \]  

where \( E_n \) is the initial error at the \( n \)th assimilation cycle and \( E_{\tau,n} \) represent the first guess value for the next step in the data assimilation cycle, \( n+1 \). The estimated variance at time step \( n+1 \) is obtained by adding the inverse of the variance of the "first guess", \( E_{\tau,n} \) and the initial error \( E_0 \) which gives

\[ E_{n+1}^{-1} = E_0^{-1} + E_{\tau,n}^{-1} \]  

By assuming that

\[ \varepsilon_n = \frac{E_n}{E_0} = \varepsilon \text{ for } \lim_{n \to \infty} \varepsilon_n = \lim_{n \to \infty} \varepsilon_{n+1} \]  

and inserting this in (3) and (4) we can finally express the relative reduction in the variance \( \varepsilon \) as a function of the model accuracy \( S \), the analysis error \( E_0 \) and the data assimilation frequency, \( \tau \). Figure 5 shows how the initial error at 500 mb height varies as a function of these three parameters. Errors are for convenience expressed in RMS and not in variance.

Using data from operational forecasts at ECMWF for the winter 1980/1981 it is found that \( E_0 \) is about (20)^2 m^2 and the model error source rate \( S \) is about (15)^2 m^2 day\(^{-1}\). For a data assimilation frequency of 12 hours, this gives an initial error of 13 m. Increasing the data assimilation frequency to 3 hours will reduce this further to 10 m. It is interesting to note the very substantial effect which is obtained by increasing the model accuracy and the data assimilation frequency. For instance, a reduction of the model error by a factor of 3, which perhaps may hardly be feasible, and a reduction of the data assimilation frequency from 12 hours to 3 hours will provide a better initial state even if the analysis error (or observation error) doubles.
The result of this exercise should not be taken literally but merely as an indication of the importance of accurate models and high data assimilation frequency in reducing the initial error.

V. CONCLUSIONS

The preliminary results so far obtained from the global weather experiments are most encouraging, and we may have confidence that the very ambitious objectives which were set up for the experiment will be met. This study which has been mainly concerned with network evaluation and observing system experiments at ECMWF has demonstrated the importance of the FGGE Observing Systems. It has been found in these studies that both aircraft data and satellite observations have a considerable impact on numerical weather prediction, at least for the large-scale flow. It is necessary to provide the basic global observing system which for practical and economic reasons to a large extent must be based on satellite observations, buoys and automatic stations. In addition to this, efforts should be concentrated on eliminating various data-void areas in active meteorological regions such as the North Pacific and Central North Atlantic with supporting high quality observations. In these areas aircraft data are of particular importance.

In view of the problems to obtain good quality data in the lower part of the troposphere by satellite soundings, it appears that upper air soundings should be concentrated for these levels. Hereby less expensive radiosonde equipment could be used. The importance of high quality models and high data assimilation frequency is stressed.

REFERENCES


Leith, C.E. 1980: Private communication

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Wind (ms⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>level radiosonde (mb)</td>
<td>TIROS-N microwave</td>
</tr>
<tr>
<td></td>
<td>clear partly cloudy</td>
</tr>
<tr>
<td>10</td>
<td>4.5</td>
</tr>
<tr>
<td>20</td>
<td>3.8</td>
</tr>
<tr>
<td>30</td>
<td>3.2</td>
</tr>
<tr>
<td>50</td>
<td>2.7</td>
</tr>
<tr>
<td>70</td>
<td>2.3</td>
</tr>
<tr>
<td>100</td>
<td>2.1</td>
</tr>
<tr>
<td>150</td>
<td>2.1</td>
</tr>
<tr>
<td>200</td>
<td>2.0</td>
</tr>
<tr>
<td>250</td>
<td>1.8</td>
</tr>
<tr>
<td>300</td>
<td>1.6</td>
</tr>
<tr>
<td>400</td>
<td>1.5</td>
</tr>
<tr>
<td>500</td>
<td>1.2</td>
</tr>
<tr>
<td>700</td>
<td>1.1</td>
</tr>
<tr>
<td>850</td>
<td>1.1</td>
</tr>
<tr>
<td>1000</td>
<td>-</td>
</tr>
</tbody>
</table>

Sea surface pressure: SYNOP/SHIP 1.0 mb; buoy 1.0 mb
COLBA/DROPWINDSONDE/TWOS-NAVAID observation errors are calculated from the level II-B quality information. Temperatures given are layer means.

Table 1: Estimated observation errors for different observing systems.
<table>
<thead>
<tr>
<th>System A.</th>
<th>All FGGE data</th>
<th>System F + (ACFT+SATEM+SATOBI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System B.</td>
<td>FGGE-ACFT</td>
<td>System F + (SATEM+SATOBI)</td>
</tr>
<tr>
<td>System C.</td>
<td>FGGE-(SATEM+SATOBI)</td>
<td>System F + ACFT</td>
</tr>
<tr>
<td>System D.</td>
<td>FGGE-(ACFT+SATEM)</td>
<td>System F + SATOB</td>
</tr>
<tr>
<td>System E.</td>
<td>FGGE-(ACFT+SATOBI)</td>
<td>System F + SATEM</td>
</tr>
<tr>
<td>System F.</td>
<td>FGGE-(ACFT+SATEM+SATOBI)</td>
<td>System F</td>
</tr>
<tr>
<td>System G.</td>
<td>FGGE-(TEMP+PILOT)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2:** Observing systems used in the prediction experiments.

The second row uses the full FGGE data set as the reference system. The third row uses system F, as the reference system. In system F we have removed all aircraft - and satellite data.
<table>
<thead>
<tr>
<th>Observing System</th>
<th>Predictability in Days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80% Good Forecasts</td>
</tr>
<tr>
<td>System A</td>
<td>All FGGE data</td>
</tr>
<tr>
<td>System B</td>
<td>FGGE-ACFT</td>
</tr>
<tr>
<td>System C</td>
<td>FGGE-(SATEM+SATOB)</td>
</tr>
<tr>
<td>System D</td>
<td>FGGE-(ACFT+SATEM)</td>
</tr>
<tr>
<td>System E</td>
<td>FGGE-(ACFT+SATOB)</td>
</tr>
<tr>
<td>System F</td>
<td>FGGE-(ACFT+SATOB</td>
</tr>
<tr>
<td></td>
<td>+SATEM)</td>
</tr>
<tr>
<td>System G</td>
<td>FGGE-(TEMP+PILOT)</td>
</tr>
</tbody>
</table>

Table 3a: Predictive skill in days at the Northern Hemisphere for 7 different observing systems. The predictability is expressed as an average for all standard levels, 1000 mb - 200 mb, and for the area 20°N-82.5°N. The figures represent the average value for 4 cases: 00Z, 10,11,13 and 16 November 1979. For further information see text.
<table>
<thead>
<tr>
<th>Observing System</th>
<th>Predictability in Days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80% Good Forecasts</td>
</tr>
<tr>
<td>System A</td>
<td><strong>All FGGE data</strong></td>
</tr>
<tr>
<td>System B</td>
<td><strong>FGGE-ACFT</strong></td>
</tr>
<tr>
<td>System C</td>
<td><strong>FGGE-(SATEM+SATOBO)</strong></td>
</tr>
<tr>
<td>System D</td>
<td><strong>FGGE+(ACFT+SATEM)</strong></td>
</tr>
<tr>
<td>System E</td>
<td><strong>FGGE-(ACFT+SATOBO)</strong></td>
</tr>
<tr>
<td>System F</td>
<td><strong>FGGE-(ACFT+SATOBO +SATEM)</strong></td>
</tr>
<tr>
<td>System G</td>
<td><strong>FGGE-(TEMP+PILOT)</strong></td>
</tr>
</tbody>
</table>

Table 3b: The same as Table 6a but for the Southern Hemisphere (20°S - 67.5°S).
Figure 1  500 mb height interpolation error for the N. Hemisphere based on a typical distribution of surface data and radiosondes (height and winds). Operational data 12Z 20 January 1982. Units are in metres.
Figure 2 500 mb height interpolation error for the N. Hemisphere based on all FGGE data 00Z 10 November 1979. Units are in metres.
Figure 3 500 mb analysis valid 00Z 11 November 1979. All FGGE data (top left); FGGE-Aircraft data (top right); FGGE-(SATEM+SATO) data (bottom left) and FGGE-(Aircraft+SATEM+SATO) data (bottom right).
Figure 4 500 mb 48 hour forecasts valid 00Z 13 November 1979
Figure 5: Error reduction due to data assimilation frequency. The figure in the matrix is the initial error in m as given by the model error, analysis error and the data assimilation frequency. For example in the top figure a model error of 15 m day$^{-1}$ and an analysis error of 20 m give an initial error of 13 m.
1. Introduction

The purpose of this study is to examine convergence properties of Machenhauer's initialization scheme. The model used is simple, so that filtering of certain nonlinear normal mode effects is easy, but the model is not low order. Only one set of data is examined. The scheme diverges in this case. The result is due to what is usually described as "the problem of small equivalent depths," except here there is a clear separation in frequencies between the gravitational and rotational linear modes.

2. Model and normal modes

The model is that of Hoskins and Bretherton (1972) with some modifications described by Errico (1982). It is adiabatic, hydrostatic, and Boussinesq, and is defined on a periodic f-plane, with fundamental wavelength $L = 15500$ km. At both $p = 0$ and $p = 200$ mb, $dp/dt = 0$.

Six orthogonal modes are resolved in the vertical. The first is barotropic and nondivergent. The remaining are baroclinic with equivalent depths $(10^4 \text{ m})/n$, for $n = 1, \ldots, 5$. Wavenumbers $0, \ldots, 21$ are resolved in both zonal and meridional directions.

The normal modes are determined from the model linearized about a state of rest, with horizontally-independent stratification $\partial \theta / \partial z = 4.5 \text{ K/km}$. For

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.
each 2-D horizontal wavenumber and each baroclinic vertical mode, there are
three normal modes. One has zero frequency and is nondivergent. The re-
maining two are gravity modes. The barotropic modes all have zero fre-
quency.

The prognostic equation for a gravity-mode coefficient, denoted $g_k$, can be written as

$$\frac{d}{dt} g_k = i\sigma_k g_k + \sum_{\ell, m} \left[ C_1 g_{k, \ell} r_m + C_2 r_k r_m + C_3 g_{k, \ell} g_m \right].$$

(1)

$\sigma$ is the frequency of the (linear) mode. The $C$'s are scale-dependent inter-
action coefficients. The $r$'s are rotational mode coefficients. Letter sub-
scripts identify particular modes.

The Machenhauer scheme attempts to solve a system $\{dg_k/dt = 0\}$, for a
set $\{k\}$, by iteration. Let a superscript denote an iteration number. Then
the scheme is

$$g_{k, I+1} - g_{k, I} = \frac{i}{\sigma_k} \frac{d}{dt} g_k$$

(2)

for the set $\{k\}$. The $r$'s remain fixed in this procedure. Here, the start-
ing iterate was $g_k^0 = 0$; i.e., a linear balance. The standard measure of
convergence is $\text{BAL} = (dg_k^I/dt)(dg_k^I/dt)^\star$.

3. Case of nonconvergence

The data examined here is generated by a diabatic version of the model
which attempts to simulate atmospheric statistics. It does reasonably
well. Day 100 is examined here. For that day, the rms vorticity (which de-
fines a Rossby number) equals 0.14.

Figure 1 describes an attempt to initialize all gravity modes given the
rotational modes. BAL is presented there, summed over all horizontal scales
for each vertical mode \( n \). After 3 nonlinear iterations, coefficients of modes of small vertical scale begin to diverge. Those of large vertical scale begin to diverge after 6 nonlinear iterations.

BAL is shown as a function of horizontal scale for the smallest vertical scale in Fig. 2. There, BAL is summed over those modes for which the horizontal wavenumber lies within a band. The band widths are all equal. The coefficients of the largest horizontal scales diverge after 7 iterations, whereas those of smaller scale diverge after 3. Thus, the coefficients which diverge first are those of modes having both horizontal and vertical scales small.

When the \( C_3 \) terms in (1) are removed, the \( g_k \) still diverge. The \( C_3 \) terms have only a minor effect in this experiment. They therefore are ignored in what follows.

4. Linear description of convergence and examination in a low-order model

When the \( C_3 \) terms in (1) are ignored, then (1) becomes linear in gravity modes. Since the rotational modes are also held fixed, the system can then be written

\[
\frac{d}{dt} \mathbf{G} = (\mathbf{S} + \mathbf{C}) \mathbf{G} + \mathbf{Q},
\]

where \( \mathbf{G} \) is a vector of \( \{g_k\} \), \( \mathbf{S} \) is a diagonal matrix of \( \{i\sigma_k\} \), \( \mathbf{Q} \) is a vector of \( \left\{ \sum_{\lambda, m} c_{2r} r_{\lambda m} \right\} \) (i.e., the quasi-geostrophic forcing term) and \( \mathbf{C} \) is a matrix with elements \( c_{1r} r_{m} \).

For (3), Machenhauer's scheme is

\[
\mathbf{G}_{i+1}^{\sim} = -\mathbf{S}^{-1}_{} \mathbf{G}_i^{\sim} + \mathbf{S}^{-1}_{} \mathbf{Q}.
\]

For \( \mathbf{G}_0^{\sim} = 0 \), the solution to (4) is
\[
G^I = - \sum_{j=0}^{I-1} (- S^{-1} C)^j S^{-1} Q. (5)
\]

If the moduli of all the eigenvalues of \( S^{-1} C \) are less than 1, then

\[
G = - (S + C)^{-1} Q \quad (6)
\]

in the limit as \( I \to \infty \); i.e.,

\[
\dot{Q} = \frac{d}{dt} C^\circ = (S + C) C^\circ + Q \quad (7)
\]

\( C \) is a 4621-squared, complex matrix. It is much too large to use with
the available computer software that performs the various operations de-
scribed in the last section. Therefore, a subset of modes was chosen to be
initialized. Only the modes having the two smallest vertical scales and
having wavelengths in both zonal and meridional directions between 1100 and
800 km were chosen. This smaller \( C \) is a 256-squared complex matrix.

The Machenhauer scheme, acting on this subset of modes does not con-
verge. Figure 3 shows the squared moduli of the 128 coefficients whose
modes have \( \sigma > 0 \). These modes behave independently of those with \( \sigma < 0 \) in
this case. Modes numbered between 65 and 128 are for the smallest-scale
vertical mode. The others are for the next larger-scale vertical mode.
Every eighth mode has the same zonal scale. Figure 3 is compared with Fig.
4 below.

For this low-order set, the matrix \( S^{-1} C \) has two eigenvalues whose
moduli are greater than 1 (valued 1.153 for the \( \sigma > 0 \) case and 1.08 for the
\( \sigma < 0 \) case). The eigenvector of the largest is presented in Fig. 4 using
the same format as in Fig. 3. The structures in both figures are similar.
The phases of the corresponding modes are also similar. Thus, there is one linear combination of gravity-mode coefficients that increases with each iteration, whereas all other linear combinations decay. If any part of the initial data projects onto this increasing combination, then after sufficiently many iterations, only this particular combination of modes will remain. In Fig. 3, many of the decaying modes are absent, and the predominance of that mode described in Fig. 4 is evident.

There is an eigenvalue of the matrix \( S + C \) which has a value \( 0.006 - 0.039i \). Thus, there is a linear combination of mode coefficients that is almost stationary when the \( r \)'s are fixed and the \( C_3 \) terms are negligible. The eigenvector corresponding to this eigenvalue is similar in structure to that of Fig. 4. Thus, the divergence of this solution to the Machenhauer scheme appears to be due to an attempt to initialize an almost stationary part of the flow. This is analogous to dividing by a near-zero value of \( \sigma \) in (2).

In the case here, there actually are no zero eigenvalues of \( S + C \). The balanced solution (7) was therefore computed directly by inverting \( S + C \).

5. Conclusion

An example is presented for which a Machenhauer balanced solution exists although Machenhauer's initialization scheme does not converge. For atmospheric conditions, convergence was shown to depend primarily on a particular matrix. This matrix describes the advection of gravity modes by the geostrophic wind field. When a particular combination of gravity modes is nearly stationary, then the Machenhauer scheme does not converge. However, it may be possible to find a Machenhauer balanced solution using some other means to invert the appropriate matrix. Standard matrix inversion techniques will not work since the matrix is too large in practice. The use of a
suitable frequency cut off (i.e., exclusion of small-scale gravity modes from the initialization) should eliminate this problem.

Fig. 1 Balance as a function of equivalent depth and iteration number

Fig. 2 Balance as a function of horizontal scale for 1.7 km equivalent depth

Fig. 3 Balance at iteration 14 for individual modes of low-order set

Fig. 4 Relative amplitude of individual modes of low-order set for unstable solution
1. Introduction

The theory of sequential estimation for stochastic-dynamic systems is a convenient framework for developing new data assimilation schemes, and for studying and improving current ones.

First, we outline sequential estimation, highlighting the Kalman-Bucy (KB) filter, and we apply the KB filter to assimilate data into a shallow-water equations model. Then we combine the KB filter with variational normal mode initialization, to produce an "optimal" combined data assimilation-initialization scheme. Finally, we indicate how the KB filter differs from widely used "optimal interpolation" schemes, and we close with some remarks concerning further applications of sequential estimation.

2. Sequential Estimation and the Kalman-Bucy Filter

Suppose a discrete, linearized numerical weather prediction (NWP) model is represented by the system

\[ \mathbf{x}_k = \mathbf{A} \mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \quad k = 1, 2, 3, \ldots \]

where \( k \) is the time index; we discuss fully nonlinear models in Section 5. To express the fact that NWP models are imperfect, we assume that the true atmospheric state, \( \mathbf{x}_k^* \), evolves according to
\[ w^t_k = w_{k-1}^t + b^t_{k-1}, \ k = 1,2,3,... \] (2)

The system noise \( \{b^t_k\} \) is assumed to be a white-noise sequence, with mean zero and covariance matrix \( Q_k \):

\[ E_b = 0, \quad E(b^t_k)(b^t_k)^T = Q_k \delta_{kk} \] (3a,b)

Here the symbol \( E \) is the expectation operator, \( (\ )^T \) is the transpose of \( (\ ) \), \( \delta_{kl} = 0 \) for \( k \neq l \), and \( \delta_{kk} = 1 \). Both \( w^t_k \) and \( b^t_k \) are n-vectors, \( n \) being the number of degrees of freedom of the model, and \( \forall_k \) is an \( n \times n \) matrix.

Observations \( w^o_k \) of the atmospheric state are made at discrete intervals, and are modeled by the equation

\[ w^o_k = H_k w^t_k + b^o_k \] (4)

Observational errors are represented by the white-noise sequence \( \{b^o_k\} \), and it is assumed that \( b^o_k \) and \( b^t_k \) are uncorrelated:

\[ E_b = 0, \quad E(b^o_k)(b^o_k)^T = R_k \delta_{kk}, \quad E(b^o_k)(b^t_k)^T = 0. \] (5a,b,c)

The observation vector \( w^o_k \) has dimension \( p_k \), and \( H_k \) is a \( p_k \times n \) matrix; typically \( p_k \ll n \) in NWP. The matrix \( H_k \) accounts for interpolation between model grid points and observation locations, and for any necessary conversion between observed variables and state variables (e.g., between satellite radiances and temperatures).

A data assimilation scheme is said to be linear and unbiased if the analysis field \( w^a_k \) is a linear combination of the forecast field \( w^f_k \).
and of the observations \( w_k^O \), and if \( E w_k^a = E w_k^f \). All such schemes may be written in the form

\[
\begin{align*}
    w_k^f &= \psi_k^{-1} w_k^a, \\
    w_k^a &= w_k^f + K_k(w_k^O - H_k w_k^f).
\end{align*}
\] (6a, 6b)

The \( n \times p_k \) gain matrix \( K_k \) specifies the linear combination: the assimilation scheme is completely determined by choosing a gain matrix sequence. Notice that all linear unbiased assimilation schemes are sequential, in that the only observations processed at a given time are the current ones.

Given the stochastic-dynamic system (2-5), one can actually determine how well a given linear unbiased data assimilation scheme performs, i.e., one can measure the success with which different gain matrix sequences reduce forecast error. It follows from Eqs. (2-6) that the forecast and analysis error covariance matrices, defined by

\[
\begin{align*}
    p_{k}^{f,a} &= E(w_w^f, a - w_k^a)(w_w^f, a - w_k^a)^T, \\
    p_{k}^{f} &= \psi_k^{-1} p_{k-1}^{f} \psi_k^{-1} + Q_k, \\
    p_{k}^{a} &= (I-K_k H_k)^T p_k^a (I-K_k H_k)^T + K_k R_k K_k^T.
\end{align*}
\] (7a,b, 8a, 8b)

To assess the performance of the assimilation scheme, one has only to advance Eqs. (8) along with Eqs. (6). The analysis error variances, which are the primary measures of performance, appear along the main diagonal of \( P_k^a \).

One would like to choose gain matrices which render the state estimates \( \{w_k^f, a\} \) optimal in some sense. Minimizing the analysis error variances with respect to the elements of \( K_k \) yields the Kalman-Bucy gain matrix,
equations (6,8,9) comprise the Kalman-Bucy filter. An equivalent defining property of the KB gain matrix is that it minimizes every quadratic functional $\eta_k(A)$ of the analysis error;

$$\eta_k = E((\mathbf{w}_k^a - \mathbf{w}_k^t)^T A(\mathbf{w}_k^a - \mathbf{w}_k^t))$$

is minimized for all positive semidefinite weighting matrices $A$. In this very broad sense, the KB filter is the optimal data assimilation scheme for the stochastic-dynamic system (2-5). See Ghil et al. (1981) and Cohn (1982) for references and further discussion of the KB filter.

3. **Application to a Barotropic Model**

We have applied the KB filter to a linear, spatially one-dimensional version of the shallow-water equations:

\[
\begin{align*}
\mathbf{u}_t + \mathbf{u}_x + \phi_x - f\mathbf{v} &= 0, \\
\mathbf{v}_t + \mathbf{u}_x + fu &= 0, \\
\phi_t + U\phi_x + \phi u_x - fU\mathbf{v} &= 0.
\end{align*}
\]

The coordinate $x$ points eastward, $u$ and $v$ are perturbation velocities, eastward and northward, and $\phi$ is the perturbation geopotential. The mean zonal current is $U = 20$ m/s, the mean geopotential is $\phi = 3 \times 10^4$ m$^2$/s$^2$, and the Coriolis parameter is $f = 10^{-4}$ sec$^{-1}$. The forecast model (6a) was obtained by a finite-difference approximation of (11); the components of $w^f_k$ are the values of $(u,v,\phi)$ at a space-time grid $(j\Delta x,k\Delta t)$, and the transition matrix $T_k$ is constant in time.
We have studied an idealized situation corresponding to the conventional radiosonde network, in which no observations are available over the "ocean" and complete observations of wind and geopotential are available over "land" every 12h; ocean and land alternate in the \( x \)-direction, each occupying half of the periodic computational domain. Experiments using other observing patterns, including satellite observations, are reported in Ghil et al. (1982). See Ghil et al. (1981) and Cohn (1982) for assumptions concerning the observational noise covariance \( R_k \), the system noise covariance \( Q_k \) and the initial error covariance \( P_0^a \).

Results of the "radiosonde network" experiments are shown in Fig. 1. The curves are marked \( U, V, P \) and \( E \) for the expected rms error in the estimation of \( u, v, \phi \) and the total energy, respectively. These results are obtained from appropriate diagonal elements of \( P_k^f/a \), and do not depend on the particular choice of initial state \( w_0^a \).

Over the data-dense land (Fig. 1a), the error immediately drops below the level of observational error, at the first observation time. It increases again in between successive observation times, due to the system noise \( b_k^t \) and due to advection of error from over the ocean.

Over the data-sparse ocean (Fig. 1b), the decrease of error at observation times is much smaller. This decrease is strictly due to the filter's spreading of information just received over land to adjacent ocean points. The increase of error in between observation times is much milder than over land: the gradual advection of information from over land partially compensates for the local system noise.

In Fig. 2 we show the evolution of the estimated geopotential field \( \phi^f/a(j\Delta x,k\Delta t) \) at three grid points, denoted by HA, NY and SF, for Hawaii, a mid-ocean point, New York and San Francisco, the easternmost and westernmost observation points, respectively. Notice that the occasional jumps at observation times at SF are larger than those at NY, since the forecast at SF has larger errors than the forecast at NY:
Figure 1

Figure 2
SF is downwind from the data-sparse ocean, while NY is downwind from data-dense land. The jumps at HA are quite small: the weighting coefficients of the KB gain matrix decrease rapidly with distance from the observations.

The overall picture in Fig. 2 is that of slowly evolving, large amplitude Rossby waves, upon which are superimposed smaller amplitude, rapidly evolving inertia-gravity waves. Due to the inertia-gravity component of the observations, the state estimate $\mathbf{w}_k^f$ does not evolve within the discrete model’s slow subspace. In order to remove the inertia-gravity component of the state estimate, we modify the KB filter as follows.

4. The Modified Kalman-Bucy Filter

As in the case of the standard KB filter, we require the modified KB filter to minimize the error functional (10), but now subject to the constraint that

$$\mathbf{w}_k^g \in R, \ k = 1,2,3,..., \quad (12)$$

where $R$ denotes the discrete model’s slow subspace. It is assumed that

$$\mathbf{w}_0^g \in R, \quad (13)$$

i.e., that initialization has been performed at the outset.

The solution of this problem (Cohn, 1982) is to take for the gain matrix

$$K_k = K_{kKB} = \Pi K_k^{KB}, \quad (14)$$
where $\Pi$ is the so-called A-orthogonal projection matrix onto $R$, defined by

$$\text{Range } \Pi = R, \quad \Pi^2 = \Pi, \quad (\Pi\Pi)^T = \Pi \Pi \ . \quad (15a,b,c)$$

The **modified KB filter** is the data assimilation scheme (6,8,14) based on the choice of gain matrix $K_{KB}$. Notice that $\Pi$, and therefore $K_{KB}$, actually depends on the weighting matrix $A$: one must now choose the error functional to be minimized.

For any given choice of $A$, the modified KB filter also has the property that it minimizes the functional

$$\tau_k = E(\bar{w}_k^a - \bar{w}_k^a)^T A(\bar{w}_k^a - \bar{w}_k^a) \ , \quad (16)$$

subject to the constraint (12), where $\bar{w}_k^a$ denotes the analyzed field that would be produced by using the standard KB filter at time $k$. In fact, we have

$$\bar{w}_k^a = \Pi \bar{w}_k^a \ . \quad (17)$$

Thus, the modified KB filter combines the standard KB filter with variational normal mode initialization (Daley, 1978; Phillips, 1981; Tribbia, 1982): $\bar{w}_k^a$ is an objective analysis, $\bar{w}_k^a$ is the initialized field, and the elements of $A$ are the variational weights. We emphasize, though, that the modified KB filter minimizes not only the A-distance (16) between the final, initialized field $\bar{w}_k^a$ and the "analyzed" field $\bar{w}_k^a$, but also the A-distance (10) between $\bar{w}_k^a$ and the true field $\bar{w}_k^t$, which is a measure of the actual analysis error.

For comparison with Fig. 2, we show in Fig. 3 corresponding results of a run using the modified KB filter based on a diagonal weighting matrix $A$, for which (10) represents the total energy of the
Figure 3

Figure 4
analysis error. The inertia-gravity waves have been completely removed.

Figures 4a,b, for comparison with Figs. 1a,b, show the expected rms errors for the run with the modified .KB filter. Inertia-gravity waves have been removed at the expense of only a slight increase in estimation error. The largest increase is in the variable \( u \), which is particularly sensitive because of the lack of a pressure-gradient term to balance the Coriolis term in Eq. (11b).

5. Further Applications of Sequential Estimation

Optimal interpolation. It was pointed out in Sec. 2 that one can study the performance of linear unbiased data assimilation methods by monitoring the covariance matrices \( P_k^{a} \), Eqs. (8a,b). "Optimal interpolation" (OI) schemes, in particular, can be written in the form (6b). The OI gain matrix, like the KB gain matrix, is derived by minimizing the variance of the analysis error in each state variable. The difference is that OI is based on a \textit{prescribed} forecast error covariance \( S_f \), rather than on the \textit{true} covariance \( P_k^{f} \), which results from assuming an underlying stochastic-dynamic atmospheric model, Eq. (2).

Thus, the OI gain matrix is given by

\[
K_k = K_k^{OI} = S_f^{k} H_k^T \left( H_k S_f^{k} H_k^T + R_k \right)^{-1};
\]

cf. Eq. (9).

OI schemes differ from each other in their formulation of the covariance matrix \( S_k^{f} \). In the operational schemes at NMC (Bergman, 1979; McPherson \textit{et al.}, 1979) and at ECMWF (Lorenc, 1981), \( S_k^{f} \) is decomposed into a \textit{time-independent correlation matrix} \( C \) and a \textit{time-dependent diagonal variance matrix} \( D_k^{f} = \text{diag} S_k^{f} \).
The correlation matrix is based on assumptions of homogeneity, isotropy and Gaussianity for mass-mass correlations, and geostrophy for mass-wind and wind-wind correlations. The correlation matrix is block-diagonal, rather than full, due to the local nature of the operational schemes: at NMC, updates use a fixed number of nearby observations, while updates at ECMWF are performed in "analysis volumes".

The variance matrix \( D_k^f \) evolves according to

\[
D_k^f = D_k^a - \tau + D ;
\]  

(19b)

here the diagonal matrix \( D \) contains prescribed growth rates of the forecast error variances and \( \tau \) is the length of the assimilation cycle. The matrix \( D_k^a \) of analysis error variances is given by

\[
D_k^a = \text{diag } S_k^a , \quad (19c)
\]

\[
S_k^a = (I-K_k H_k)S_k^f (I-K_k H_k)^T + K_k R_k K_k^T ; \quad (19d)
\]

\( S_k^a \) is the OI estimate of the analysis error covariance resulting from use of the gain matrix \( K_k = K_k^{OI} \). Equations (19) can be regarded as an approximation to the evolution (8) of the true covariance \( P_k^{f,a} \).

Cohn et al. (1981) have studied the performance of the OI filter given by Eqs. (6,18,19) for the shallow-water model described in Sec. 3, using a variety of choices for the matrices \( C \) and \( D \). The results show that, while the simplifying assumptions made in OI increase computational efficiency, they have undesirable effects on the results of the assimilation. In particular, homogeneous, isotropic forecast error correlations lead to poor analyses near boundaries separating data-dense and data-sparse areas. This problem can be
partly alleviated by using in the matrix $D$ forecast error growth rates which depend on data density: we have already seen (Figs. 1, 4) that as a result of advected error, growth rates for an optimal filter must depend very strongly on data density. Proper initialization is also a partial cure for this boundary problem.

**Adaptive filtering.** Model error and observational error are important quantities in the statistical weighting of data during linear, unbiased estimation. In Sec. 2, the system noise covariance $Q_k$ and the observational noise covariance $R_k$ were assumed to be known a priori. This does not reflect the actual state of affairs in NWP. Fortunately, $Q_k$ and $R_k$ can be determined adaptively, i.e., during the assimilation process itself. Biases in the system and in the observations, assumed zero in Eqs. (3a, 5a), can also be determined adaptively.

There are a number of approaches to adaptive estimation (Chin, 1979). An approach that appears promising for NWP is that of Bélanger (1974). We assume for notational simplicity that $Q_k$ and $R_k$ are time-independent: $Q_k = Q$ and $R_k = R$.

Suppose that $Q$ and $R$ are represented as linear combinations of some judiciously chosen matrices $Q^i$, $R^j$,

$$Q = \sum_{i=1}^{M} \alpha_i Q^i, \quad R = \sum_{j=1}^{N} \beta_j R^j; \quad (20a,b)$$

for instance, $Q_1$ and $Q_2$ can represent the geostrophic and ageostrophic parts of the system noise, respectively, with $\alpha_1$ and $\alpha_2$ some amplitude factors. The problem, then, is to estimate the unknown coefficients $\alpha_i$, $\beta_j$. Given Eqs. (20), it follows from Eqs. (2-6) that

$$E[r_k r_k^T] = \sum_{i=1}^{M} \alpha_i F_i(k,\ell) + \sum_{j=1}^{N} \beta_j G_j(k,\ell), \quad (21)$$

where
\[ \varepsilon_k = \mathbf{x}_k^O - H_k \mathbf{x}_k^f \]  

(22)

is the observed-minus-forecast residual, and where \( P_1 \) and \( G_j \) are known, recursively defined, \( P_k \times P_{k-2} \) matrices. The lag-covariance matrix on the left-hand side of Eq. (21) can be estimated from the sample output of the assimilation scheme. The parameters \( \alpha_i \) and \( \beta_j \) are determined by a least-squares fit to Eq (21); \( Q \) and \( R \) are then given by Eq. (20).

**Nonlinear filtering.** In sequential estimation for a realistic, nonlinear forecast model, Eq. (6a) is replaced by

\[ \mathbf{w}_k^f = f(\mathbf{w}_{k-1}) ; \]  

(23a)

the update procedure is still linear:

\[ \mathbf{w}_k^a = \mathbf{w}_k^f + K_k(\mathbf{w}_k^O - H_k \mathbf{w}_k^f) . \]  

(23b)

The underlying stochastic-dynamic atmospheric model (2) is replaced by

\[ \mathbf{w}_k^t = f(\mathbf{w}_{k-1}) + \mathbf{b}_{k-1}^t , \]  

(24)

while assumptions (3-5) remain in effect.

The main difficulty in nonlinear filtering is that the forecast error covariance matrix \( P_k^f \) now depends on all the higher moments of the analysis error at time \( k-1 \), not just on the second moments \( P_k^a \) as in Eq. (8a). However, if we define \( \Psi_k \) to be the Jacobian matrix

\[ (\Psi_k)^{1,j} = \left. \frac{\partial f^i(\mathbf{w})}{\partial w^j} \right|_{\mathbf{w} = \mathbf{w}_k^a} , \]  

(25)
where \( \bar{w}_k \) is a reference state, then the error evolution equation (8a) is still correct to first order in \( \bar{w}_k - \bar{w}_k^f \). Consequently, the KB gain matrix (9) is a first-order approximation to the optimal gain matrix for our nonlinear system, and Eq. (8b) gives the corresponding analysis error covariance. That is, Eqs. (23, 25, 8, 9) comprise a nonlinear filter which is optimal to first order.

There is, of course, substantial leeway in the choice of reference state \( \bar{w}_k \). The filter corresponding to the obvious choice \( \bar{w}_k = \bar{w}_k^f \), for which the Jacobian (25) must be computed every time step, is known as the extended Kalman filter (EKF). The EKF has been successful in a wide variety of engineering applications, and performs well when the system noise \( Q_k \) and initial uncertainty \( P_0 \) are small. Jazwinski (1970) gives a detailed treatment of nonlinear filtering, including higher-order filters and the case of nonlinear observation models.

It is probably not necessary in NWP to compute the Jacobian every time step. Rather, advancing the forecast error covariance by linear dynamics over much longer intervals is probably adequate. We expect simplified versions of a fully optimal adaptive nonlinear filter to be efficiently computable for NWP. The challenge is to find the best compromise between optimality and computability.

Acknowledgements. We wish to thank E. Isaacson and the participants of the Seminar for many stimulating discussions. This work is supported by the National Aeronautics and Space Administration through grants NSG-5130 (M.G.) and NSG-5034 (S.C.), monitored by the Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center.

References


Ghil, M., S. Cohn, and A. Dalcher, 1982. Applications of sequential estimation to data assimilation. Proc. Study Conf. on Large-Scale Oceanographic Experiments in the World Climate Research Program, Tokyo, 10-21 May 1982, National Center for Atmospheric Research, Boulder, Colorado 80307, pp. 67-86.


This paper gives an account of some current work of the Data Assimilation and FGGE groups at ECMWF. Research and development work covers a variety of problems in most aspects of operational data assimilation.

The topics to be treated and the people involved are listed below:

- Data monitoring (D/A group)
- Resolution of an analysis system (Lönnberg, Hollingsworth)
- Determination of structure functions (Lönnberg, Hollingsworth)
- The implementation of a diabatic initialisation (Wergen)
- Tropical Analysis 1. Dynamical Theory (Hollingsworth)
  2. Idealised Experiments (Cats)

DATA MONITORING

We find a growing need to devote more resources to efforts to monitor data quality. Many of the problems are well known - radiosonde biases in wind as well as height, mislocated ships, stations that do not follow standard practices. We outline our current activity in this area as well as plans for future work. Every operational centre has its own techniques for dealing with the problems. We feel that there is much to be gained by a more active exchange of information between operational centres on such matters.

Finally, we feel that there is a need to re-examine some of the standard procedures themselves, particularly near mountains. We present an example of a defective analysis resulting from good data, correctly reported. The problem is that the standard reporting practices are unreasonable.

ANALYSIS RESOLUTION

We show an example of an inferior one-day forecast. A study of the initial analysis leads to the conclusion that the problem arose because the system to be analysed
was of such small scale as to be well below the resolution of the analysis system. We devised a simple measure of the resolution of an analysis system and explored the resolving power of the analysis as a function of observational error and the scale of the structure function. The theoretical results gave good agreement with the experimental results. The significance of these results for analysis of the thermal wind field is briefly discussed.

STRUCTURE FUNCTIONS

We present some preliminary results on the determination of structure functions for the forecast error of the ECMWF model. The results are encouraging in that these empirically determined structure functions offer a number of advantages over the currently operational structure functions. Specifically, they offer us the potential for a higher resolution analysis, and secondly, they will eliminate a number of shortcomings in the wind analysis which arise from the use of Gaussian structure functions.

DIABATIC INITIALISATION

Work has been completed on a procedure to incorporate diabatic terms in the non-linear normal mode initialisation. The main problem has been to prevent the excitation of intense small scale convective systems. This was done by using a forcing term which is the same at each iteration and which is heavily smoothed. The forcing is determined as the time average of the forcing in a two-hour forecast from uninitialised data. The spatial smoothing is accomplished by eliminating from the forcing field all its projections on gravity waves with periods shorter than eight hours. This cut-off was chosen so as to include the main thermal tides. Examples of the results are shown.

TROPICAL ANALYSIS

We extend the work of Gill on steady forced tropical circulations to a primitive equation formulation. We show that this can improve some important features of Gill's results. The steady state results can be extended to time dependent problems and we show some examples where they provide a useful interpretation of short range forecast errors. An important implication of the steady state results is that a tropical analysis scheme must be able to analyse both Rossby wave and (forced) gravity wave motion. Some deficiencies of current formulations in this regard are illustrated.

Acknowledgments

I am grateful to my colleagues at ECMWF for the use of their unpublished results.
I will be concerned here with objective analysis of the wind field in the tropics: the material will not be concerned directly with the initialization problem nor with analysis of the mass field. The title perhaps is misleading since I will not suggest that the analysis variables be, specifically, streamfunction and velocity potential. What I will suggest is that objective analysis techniques in the tropics need to accommodate both the rotational and irrotational components, and that estimates of the irrotational component are best obtained from non-conventional observational sources.

Examination of the behavior of a very sophisticated and finely-tuned objective analysis scheme, that of the ECMWF (Lorenc, 1981), operating with the FGGE data base, has convinced me (Julian, 1981a,b,c) that (i) current analysis schemes do not consistently and predictably provide proper structure of the wind field in the vicinity of strong convergence or divergence; and (ii) the important areas of the tropics with these strong divergence fields are those associated with organized deep convective activity. We have available, but do not make use of, information on the location and intensity of this convective activity.

There is abundant evidence in the literature that to a very good first approximation a universal vertical divergence profile exists in the tropics, the amplitude and sign of which is given by the strength of the convection (e.g., Gray et al., 1982; McBride and Gray, 1980; Frank, 1978; U.S. GATE Committee, 1977). Since the satellite window-channel, infrared radiometer data provides us with a quantitative estimate of cloud top altitude, I am suggesting that these data be used as proxy data to estimate the irrotational component of the wind field. Similar suggestions have been made by Chang (1975), to balance the synoptic-scale vorticity imbalance known to exist in the tropics; and by Sumi et al. (1979), to assist in an objective analysis of the tropical wind field.

Figure 1 is an example which illustrates point (i) above. It is a FGGE III-b 250 mb map (ECMWF) for 6 May 1979, 00 GMT, showing both the analyzed wind field and the II-b data available to the analysis scheme. Note that rawin observations from stations 91434, 91317, and 91334 and omega sounding MANUS have either been rejected or given near zero weight by the analysis scheme. In addition, aircraft wind TTRS5 has been similarly ignored. I have added with heavy solid isopleths areas with cloud infrared data colder than \(-35^\circ\text{C}\), which indicates areas with deep convection occurring. The expected outflow from these areas is consistent with the observed, but ignored, winds. The combined data selection and optimum interpolation schemes have produced
an analysis here which clearly can be improved upon. Other cases examined but not shown here (but see, e.g., case 1 in Julian, 1981a) often show a form of aliasing in which the strong divergent component indicated by the observations is distorted into the largely rotational, analyzed component. Presumably, this is caused by the use of non-divergent structure functions in the statistical analysis scheme. As stated above, the quality of the analyses is neither consistent or predictable. In Fig. 2, I show an example of, again, a Level III-b ECMWF analysis in which the irrotational component as analyzed is in good agreement with that expected from the location and strength of the deep convection. The maps, 200 mb above and 850 mb below, are for 10 May, 00 GMT. Note also that the divergent component is also in expected agreement with that expected in the cloud-free areas—those with infrared temperatures above -10°C. There are some features in this example which indicate that some improvement might be expected. For example, the course/sink region in the irrotational wind at 12N, 82E is not centered over the tropical disturbance located there (cloud IR temperatures below -50°C). And there is an analyzed divergent region at 200 mb located south of Java and Sumatra which is not supported by the presence of any convective activity.
The algorithm I have at present for analyzing the irrotational component proceeds as follows: the infrared radiometer data which is arranged over and on a 2.5° latitude-longitude grid is scaled to divergence using a non-linear relationship which assigns significant divergence/convergence to those areas with equivalent temperatures less than 235 K/greater than 269K. Boundary conditions on the velocity potential at ±30° latitude are obtained from the conventional analysis. The appropriate Laplace equation for velocity potential is then solved and the irrotational components obtained. I have found by experimentation that proceeding vis the divergence is preferable to simply scaling the radiometer data directly to velocity potential. In addition, experimentation also has suggested that the use of a high wave number emphasis digital filter applied to the radiometer data produces a velocity potential having better detail at scales less than 10°.
Figure 3
A sample of the velocity potential produced by this algorithm compared with that obtained from a conventional (but single level, wind only) optimum interpolation scheme is shown in Figs. 3 and 4. The conventional analysis for 3 Sept 1975, 00 GMT, during Data Systems Test 5, is shown (3, top) by a streamline analysis with the available data also added. Figure 3, bottom, simply shows where the IR data was less than $-28^\circ C$ (2 digits) or greater than $-3$ (single digit). In Fig. 4, top, the velocity potential field from the conventional analysis scheme can be compared with the same variable obtained from the IR data 4, bottom. Comparison shows good agreement on the largest scales and in regions of adequate conventional data but substantial differences in data-poor regions.

Current work involves the determination of the best method of incorporating the irrotational component derived from the IR data into the overall statistical analysis scheme.

References


The bounded derivative method (BDM) proposed by Kreiss (1979) can be applicable to formulate an initialization scheme for atmospheric prediction models. Let us consider an ordinary differential equation in the form
\[ \frac{dW}{dt} = L(t) W + F(t), \]  
where \( W \) is an \( n \)-component vector, \( L(t) \) is a linear \( n \times n \) matrix which depends only on time \( t \), and \( F(t) \) denotes a time dependent forcing term. We assume that matrix \( L \) has \( n \) pure imaginary eigenvalues, \( \nu \), which are split into the two classes
\[ C_1 = \{ \nu_1, \ldots, \nu_{\frac{\nu}{2}} \}, \quad \text{(first class)} \]
\[ C_2 = \{ \nu_{\frac{\nu}{2}+1}, \ldots, \nu_n \}, \quad \text{(second class)} \]
where \( \nu_j = O(\varepsilon^{-1}) \) for \( \nu_j \) in \( C_1 \) and \( \nu_j = O(1) \) for \( \nu_j \) in \( C_2 \). Here, \( \varepsilon = O(10^{-1}) \). The frequencies \( \nu_j \) in \( C_1 \) are one order of magnitude larger than those in \( C_2 \). Hence we call motions of \( C_1 \) as fast modes and those of \( C_2 \) as slow modes.

The BDM is based on the following observation: If \( W(t) \) varies slowly, then
\[ \frac{d^m W}{dt^m} = O(1) \quad \text{for} \quad m = 0, 1, \ldots, M, \]  
where \( M > 1 \) is some suitable number.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.
When this observation is applied at $t = 0$, the following scheme of initialization is derived: Choose the initial value $W(0) = W_0$ such that

$$\frac{d^m W}{dt^m} = 0(1) \quad \text{at} \quad t = 0 \quad \text{for} \quad m = 0, 1, \ldots, M . \quad (3)$$

The task of formulating an initialization scheme is to derive the appropriate diagnostic constraints on initial data which satisfy condition (3) and suppress the appearance of fast modes during the subsequent time integration of $W$. Bube and Ghil (1981) discussed an example of an ordinary differential equation for illustrating the elimination of fast mode solutions with the application of BDM.

Kreiss (1980) extended the application of BDM to hyperbolic partial differential equations. Browning et al. (1980) applied the BDM to shallow water equations on midlatitude and equatorial $\beta$-planes. By applying atmospheric scalings (Phillips, 1963), dimensionless scaled shallow water equations are expressed by

$$u_t + uu_x + vu_y + \varepsilon^{-1}[\phi_x - (f_0 + \varepsilon\beta y)v] = 0 , \quad (4)$$

$$v_t + uv_x + vv_y + \varepsilon^{-1}[\phi_y + (f_0 + \varepsilon\beta y)u] = 0 , \quad (5)$$

$$\phi_t + (u\phi)_x + (v\phi)_y + \varepsilon^{-2} (\phi_o - \varepsilon\Phi)(u_x + v_y)$$

$$= \varepsilon^{-1}(u\Phi_x + v\Phi_y) , \quad (6)$$

where $u$ (x-component of velocity), $v$ (y-component of velocity), $\phi$ (geopotential deviation) are dependent variables, $f_0 + \varepsilon\beta y$ denotes the Coriolis parameter, $\phi_o$ is the geopotential corresponding to the mean height of free surface, and $\Phi$ represents the geopotential corresponding to the elevation of orography. The right-hand side of (6) denotes the mountain-induced divergence effect.
The term multiplied by $e^{-1}$ or $e^{-2}$ in (4)-(6) has the magnitude of $0(10)$ or $0(10^2)$. Hence, fast time scale motions appear during their time integration if arbitrary initial conditions of $u$, $v$, and $\phi$ (even though they are of $0(1)$) are selected. In the case of midlatitude $\beta$-plane, we must, therefore, impose the conditions

$$
\phi_x - f_0 v = \varepsilon a(x,y,t),
$$
$$
\phi_y + f_0 u = \varepsilon b(x,y,t),
$$
$$
u_x + v_y = \varepsilon S^{-1}[u\phi_x + v\phi_y + \varepsilon c(x,y,t)] ,
$$

with $S(x,y) = \phi_y - \varepsilon \phi_x$, where $a$, $b$, and $c$ are of $0(1)$ and smooth. These conditions are necessary to ensure that $u_t$, $v_t$, and $\phi_t$ remain to be of order unity for a certain time integration period $T$.

The initialization problem is the task of determining the distributions of $a$, $b$, and $c$ at $t = 0$. The simplest way to ensure (7) at $t = 0$ is to assume that $a = b = c = 0$. These conditions lead to geostrophic initial conditions for the geopotential $\phi$ and the streamfunction $\psi$ as derived by Hinkelmann (1951) and the divergence is specified by the mountain effect term.

The assumption of $a$, $b$, and $c$ being all zero is too crude. Higher order constraints can be derived by assuming that $a_t$, $b_t$, and $c_t$ are of the order of unity, since these constraints ensure that $u_{tt}$, $v_{tt}$, and $\phi_{tt}$ are of the order unity at $t = 0$.

A special case of the second-order constraints leads to, in the case of midlatitude $\beta$-plane, the traditional balancing based on quasi-geostrophic theory which consists of the balance equation (Charney, 1955) relating the streamfunction to the geopotential and the barotropic version of $\omega$-equation (Phillips, 1960) for the divergence field. At this order of accuracy,
therefore, there is no advantage of the BDM. However, this approach is more
general than the use of geostrophic assumptions in order to formulate the
initialization procedure. We can proceed to deriving the third-order con-
straints by bounding the terms of $a_{tt}$, $b_{tt}$, and $c_{tt}$.

The application of BDM to the shallow-water equations on equatorial
$\beta$-plane (Browning et al., 1980) leads to a new balance equation for the geo-
potential which involves not only the streamfunction, but also some effects
of the divergence induced by orography. A numerical example, run by F.
Semazzi of the University of Nairobi, demonstrates that the use of this new
balance equation together with the mountain-induced divergence field sup-
presses effectively the generation of high-frequency gravity waves.

Kasahara (1982) extended the application of BDM to formulate an ini-
tialization procedure for a baroclinic primitive equation model with $\beta$-plane
geometry in pressure coordinates. To the degree of approximation employed,
the initialization by the BDM agrees with the classical balance procedure
with quasi-geostrophic assumptions. We demonstrated also for the same pre-
diction model that nonlinear normal mode balancing leads to an initializa-
tion scheme identical to the one derived from the BDM within the degree of
approximation. Earlier, Leith (1980) has shown a connection between the
nonlinear normal mode initialization and the classical balancing due to
quasi-geostrophic theory. The three-way connections between the BDM, the
nonlinear normal mode initialization, and the quasi-geostrophic balancing
for a simple baroclinic model are well established.

Since the BDM is free from such a restriction as quasi-geostrophy, the
method is applicable to the problem of initialization for mesoscale and
cumulus cloud models. We presented at the seminar the scale analysis of
atmospheric motions and compared the characteristics of scaled equations for large-scale, mesoscale, and cumulus-scale motions.

The basic equations of atmospheric motions, which consist of those of horizontal and vertical motions, mass continuity, and thermodynamics, are nondimensionalized according to the characteristic magnitudes of horizontal and vertical scales, horizontal and vertical particle speeds, pressure and density variations, and the time scale for the three different types of motions, namely large-scale, mesoscale, and cumulus-scale. It was noted that it is useful to subdivide the mesoscale category further into two classes of mesoscale (I) and (II).

The conclusions derived from this analysis are that the large-scale motions are quasi-horizontal, quasi-geostrophic, and hydrostatic. The mesoscale (I) motions are quasi-horizontal and hydrostatic but they are not quasi-geostrophic. The mesoscale (II) and cumulus-scale motions are not hydrostatic, but satisfy the anelastic approximation as discussed by Batchelor, Ogura, Charney, and Phillips. There are constraints on the values of the static stability of the basic state depending on the different scales of motion.

The requirement for initial data is clearly different depending on the scale of motion under consideration. The initialization procedure of the primitive equation model for mesoscale (I) motions may be formulated in the same way as that for large-scale motions using the BDM. However, for mesoscale (II) and cumulus-scale motions, accurate observations of both mass and velocity fields are necessary and the initialization procedure of the anelastic prediction model is quite different from that of the primitive equation model for large-scale and mesoscale (I) motions.
The BDM provides a general and systematic procedure to derive the diagnostic relationships under which the solutions of prediction models remain on the order of magnitude of desired motions. Although the BDM utilizes an asymptotic expansion, the emphasis here is not so much on the nature of asymptotic expansion, but on the smoothness of solutions. It is the smoothness which decides whether we can integrate the prediction system numerically for a reasonable time period (Kreiss, 1979). This emphasis on the smoothness of solutions becomes useful for deriving the practical procedure of initialization in the case of prediction equations containing forcing terms and/or the boundary conditions for a limited integration domain.

References


The FSU Global Spectral Model, Krishnamurti et al (1981), has been used to carry out a number of medium range numerical prediction experiments over the globe. The data sets for the present series of experiments are based on the FGGE/MONEX observations. The spectral model has eleven vertical levels in the sigma coordinate below 100 mb. The model resolution is 29 waves, rhomdoidal truncation. The emphasis of this study is on low latitude problems. Figure 1A illustrates a schematic outline for a new initialization procedure that is tested in a series of experiments. The scheme is designed to provide a reasonable diabatic forcing after the initialization. The satellite rain guage derived rainfall (schematically outlined in Fig. 1B) provides the convective heating rates which is built into the divergent motion field, prior to dynamic initialization. Here the adjustment of the divergent wind uses the Kuo's scheme II described in Krishnamurti et al (1980 MWR) as a framework. In the rainfree areas a method for adjustment of the humidity profile is demonstrated to provide a balance among the advective and radiative forcing. The next step is a dynamic initialization which carries out a number of steps of forward-backward integration around time zero. Subsequent to dynamic initialization no tempering with the basic variables such as temperature, pressure, horizontal wind field and vertical velocity is permissible. However an lateration of the humidity variable (which is passive during dynamic initialization, except for virtual temperature corrections) is possible. This is done to restore once again a balance among the advective and radiative processes in rainfree areas and to provide a rainfall rate elsewhere which is exactly equal to that prescribed by the satellite-rain guage measurements. The paper discusses an analysis of this scheme and its impact in 10 day forecasts. The case studies deal with formation of a tropical depression over the Northern Arabian Sea.

Adjustment of VAS Geopotential Analyses by using precedent RAOB Analyses and Quasi-geostrophic Constraints

by

John M. Lewis
Development Lab, NOAA/NESS
1225 West Dayton Street, Madison, WI 53706

1. Introduction

Radiometric data from the VISSR Atmospheric Sounder aboard the GOES series of geostationary satellites is converted to temperature using a physical, as opposed to statistical, algorithm and requires a guess. Currently, a neighboring raob analysis or 12-24 h prediction from the Limited Fine Mesh (LFM) model is used. As the frequency of asynoptic data increases through the concerted use of the polar orbiting and geostationary satellites, the guess field can hopefully be improved.

Typically more than 150 soundings are available from the Rockies to the East Coast on a 3-hourly basis from GOES-east. As envisioned, the guess will be taken from the previous analysis (3-h old) where coupling in time is enforced through the use of quasi-geostrophic constraints.

The sequence of analyses/guesses would proceed as follows:

1. The 1200 GMT analysis of geopotential from RAOB data would serve as a guess for the 1500 GMT satellite retrievals.
2. The 1500 GMT heights from the satellite would be tested for dynamical consistency with the 1200 GMT analyses. The constraint used is conservation of potential vorticity.
3. The adjusted 1500 GMT analysis is used as a guess for the 1800 GMT analysis.
4. The 1800 GMT analysis is tested for dynamical consistency with the 1500 GMT analysis.

etc.

The philosophy of this approach is based on the maxim that an analysis in the recent past is more reliable than a relatively long range forecast, especially when the analyses have been coupled temporally through dynamical constraints.

The initial testing is based on a 2-parameter filtered model with
heights specified at 250 and 700 mb. The quasi-geostrophic potential vorticities are conserved in this model and they are given by:

\[
\begin{align*}
\frac{\partial \zeta_{250}}{\partial t} &= \frac{\partial}{\partial z} \left( \frac{1}{f_0} \frac{\partial z_{250}}{\partial z} \right) + f - \lambda^2 \frac{\partial}{\partial z} \left( z_{250} - z_{700} \right) \\
\frac{\partial \zeta_{700}}{\partial t} &= \frac{\partial}{\partial z} \left( \frac{1}{f_0} \frac{\partial z_{700}}{\partial z} \right) + f + \lambda^2 \frac{\partial}{\partial z} \left( z_{250} - z_{700} \right)
\end{align*}
\]

where \( \lambda^2 \) is the static stability parameter \( \sim 1.5 \times 10^{-12} \text{m}^{-2} \). During the second stage of the analysis/guess procedure outlined above, the RAOB and satellite analyses are tested for dynamical consistency using these constraints. This step in the procedure is tested by using data on March 6, 1982, from the conventional RAOB network at 1200 GMT and the VAS analysis at 1430 GMT.
2. **Schematic View of Adjustment Procedure**

The mathematical details of the adjustment process can be found in Thompson (1969) and Lewis (1981). Least squares adjustments are made to potential vorticities at each time level subject to the constraint that the hindcast and forecast to the intermediate time are equal. Corresponding height adjustments are found by inverting the elliptic operators relating potential vorticities and heights. Four input fields are used: 700 mb heights at the early and late times, (n-1) and (n+1) respectively, and the corresponding 250 mb heights. From these input heights, Fjørtoft steering currents (Fjørtoft, 1952) are constructed at each level by spatially and temporally averaging. This is depicted at the top of Fig. 1. The potential vorticities are analyzed along the steering currents and the distribution along a particular contour (ABC) is
displayed. The amplitude of the vorticity maximum at (n+1) exceeds that at (n-1) and the measure of inconsistency is shown by the difference between hindcast and forecast, the darkened area between the curves. The maximum are also displaced at nat as indicated by $\uparrow$ $\downarrow$. As expected on heuristic grounds, the adjustment process will decrease the amplitude of the vorticity maximum at (n+1) and increase the amplitude at (n-1), by comparable amounts assuming equal weighting in the least squares process. Also, the phase will be adjusted in such a way that the maximum is moved forward at (n-1) and backward at (n+1) to compensate for the discrepancy in the forecast/hindcast. The corresponding height adjustments generally show an equal but opposite magnitude and the pattern is displaced in the direction of the current.

3. **Case Study**

Data collected on March 6, 1982, is used in this test. The visible imagery from GOES-east is shown in Fig. 2, with a large swath of clouds running from Mexico up through the Great Lakes. The cloud band was associated with a relatively strong front, but there was no active cyclogenesis in the area. Some rain showers and several thunderstorms were reported near the Gulf Coast.

Observations of geopotential at 700 and 250 mb are superimposed over the visible imagery in Fig. 2. The obvious differences between RAOB and VAS data sources are the data density and horizontal resolution. The RAOB data is uniform over the continent, but absent over the Gulf of Mexico. The average separation over the continent is 300-400 km. The VAS data is conspicuously sparse in the cloudy areas, because of the infrared sensor's sensitivity to cloud, but very dense in the cloud free areas. Data in also available over the Gulf.
The analysis approach can follow either of two paths: 1) analyze on a grid commensurate with observations from RAOB network with a resultant smoothing of the VAS data, or 2) analyze on a grid commensurate with the VAS data and introduce short wavelength noise into the RAOB data. If course 1) is followed, one must grapple with the attendant problem of boundary points interfering with the solution on the interior. In fact, there are so few grid points in the present case when a 4° latitude mesh is used, the adjustment process is incapable of giving meaningful results. Consequently, course 2) is followed using a grid interval of 1° latitude with the additional constraint that the resultant analyses are filtered to remove scales below ~600 km. The filtering is accomplished by using the scale decomposition procedures developed by Fjørtoft (1952).

The objective analyses are derived from the Cressman technique (Cressman, 1959) with distance dependent weighting of residuals (guess minus observation) where the weighting function is empirically specified. The guess field is the LFM analysis valid at 1200 GMT. This guess is used for both RAOB and VAS analyses. Since there are no rawinsonde observations over the Gulf, the RAOB analysis in this region is the LFM forecast.

Geopotential analyses with geostrophic relative vorticity superimposed are shown in Fig. 3. The progressive movement of the trough from the Texas panhandle to southeastern Oklahoma is evident by examining either the relative vorticity or potential vorticity (Fig. 4). The strength of the vorticity associated with this trough is slightly larger on the VAS analysis at 700 mb, while smaller at 250 mb.
The separation between the vorticity centers associated with this trough is excessive at 250 mb. That is, using the steering current speed at that level the displacement is 400 km, whereas the separation indicated is 550 km. Another source of discrepancy between the two analyses is the pronounced ridge over the western Gulf as indicated on the VAS analysis. Examination of colocated observations along the Texas and Louisiana Coast indicate that the rawinsonde observations are slightly higher than satellite observations at both 700 and 250 mb, i.e., there doesn't appear to be a systematic bias toward higher geopotential heights from the satellite observations where comparisons are possible.

When equal weight is given to the potential vorticity analyses from RAOB and VAS, the corresponding height adjustments (m) indicated in Fig. 5. As expected on heuristic grounds, the ridge over the Gulf is weakened on the VAS analysis while a ridge is built into the RAOB analysis in this region. Since the changes are in the same sense at 700 and 250 mb on each analysis, i.e., RAOB analyses increase at both 700 and 250 mb while VAS analyses decrease at both levels, the principal change is in the relative vorticity at each level and not the thickness.

With regard to adjustments in the trough, the 250 mb vorticity center is strengthened in the Texas panhandle and weakened over southeastern Oklahoma, in effect pulling the input vorticity maximum to the west. The corresponding adjustment on the RAOB analysis is a weakening of the center over the panhandle and a strengthening over southeastern Oklahoma.
References


RAOB

250mb

7 1

700 mb

HEIGHTS (m) AND RELATIVE VORTICITY (10^{-5}s^{-1})

VAS

250mb

7 1

700 mb

HEIGHTS (m) AND RELATIVE VORTICITY (10^{-5}s^{-1})
Figure 4

RAOB

VAS

250 mb

700 mb

700 mb

POTENTIAL VORTICITY ($10^{-5} \text{ s}^{-1}$)
ON DIRECT DATA ASSIMILATION INTO A FORECAST MODEL

by A.C. Lorenc (U.K. Meteorological Office)

1. Historical review and description of method

The advent of satellite based observing systems in the 1960s raised the possibility that near continuous observations of some atmospheric parameters would soon be available. Beginning with a paper by Charney et al (1969), research was stimulated into ways of assimilating these data into numerical forecasting systems. It can be shown, at least for linearized models which accurately represent the atmosphere, that knowledge of one parameter (e.g. wind) could be "induced" from a knowledge of a time history of another (e.g. temperature) (Davies and Turner 1977, Talagrand 1981). The method of induction proposed by Charney et al was simply to insert the data into a forward running model. In idealized cases, where the model is made to be completely accurate by assimilating pseudo-observations from an "identical twin" run of the same model, this gave encouraging results. However more realistic studies (e.g. Miyakoda et al 1978) showed that although data assimilation avoids the "spin up" problems associated with some other methods of analysis and initialization (evident for instance in the rainfall, see 4a below), care must be taken to extract the maximum information from the data while shocking the model as little as possible.

The method I shall be discussing for doing this was first used for idealized data network studies by Lorenc (1976). It was then developed for practical use during FGGE (Lyne 1980), where it was used to produce global analyses for general circulation studies during the two Special Observing Periods (Lyne et al 1982). This FGGE scheme is still in use for observing system studies. Meanwhile the method is being developed and tested for operational use at the U.K. Meteorological Office.
The basic method is a repeated interpolation and insertion of data
each timestep of a forward-running forecast model. The normal prognostic equa-
tions of the model, represented schematically by \( \vec{\psi}_{t+\Delta t} = M(\vec{\psi}_t) \), are
modified to give
\[
\vec{\psi}^*_{t+\Delta t} = M'(\vec{\psi}_t) \quad \text{and} \quad \vec{\psi}_{t+\Delta t} = \vec{\psi}^*_{t+\Delta t} + \lambda \sum_{i=0}^{n} \omega_i (\vec{\psi}^{obs}_i - \vec{\psi}^*_{i,t+\Delta t})
\]
The weights \( \omega_i \) are calculated by simple univariate optimal interpolation
(although they are not used in an optimal fashion), and \( \lambda \) is typically in the
range 0.1 to 1.0. The model equations \( M \) are modified by adding extra diffusion
and damping of high frequency and divergent model. Although the method can in
principle be used for asynoptic observations, implementations so far have grouped
the data into 6-hourly batches, and kept the weights \( \omega_i \) (but not \( \lambda \)) constant over
a 6-hour period. Refinements which are currently being tested in the Met. Office
operational trials include the addition of geostrophic wind increments calculated
from the mass field increments, and the calculation of the weights \( \omega_i \) by multi-
variate optimal interpolation.

It is perhaps worth emphasizing that the scheme differs from that
developed at GFDL (Miyakoda et al 1976), in that the interpolation is repeated
each timestep. It can thus be viewed as a successive correction method, with a
complex filter \( M' \) applied between each scan. This suggests the idea, as yet un-
tested, of using a simpler and cheaper method of calculating the weights \( \omega_i \), en-
abling them to be recalculated each step using varying radii of influence.

2. Strengths of the method

(a) It is simple in concept.

(b) It can be used for asynoptic data.

(c) It can be used for limited area high resolution or mesoscale models.
(d) It avoids the interpolation and initialization problems of schemes which have the analysis distinct from the model (see 4a and 4b).

(e) It works. The general circulation statistics calculated from the FGGE IIIa analyses compare well with those from other schemes (see 4b). A short period of FGGE IIIb analyses are currently being compared with analyses of the same data made at ECMWF and NMC. Preliminary subjective and objective comparison indicates that differences are rather minor in the northern hemisphere (see 4c).

It is too early to quote results from the operational trials now in progress, but the impression given by a small number of comparisons for the N. Atlantic area with forecasts from ECMWF analyses is that any differences in predictive skill due to the analysis are very small compared with the week to week variability in skill and not bigger than those due to differences between forecast models.

3. Weaknesses of the method

(a) It uses ad hoc rather than theoretically based methods to express our knowledge about the likely structure and scale of atmospheric motion. This means that there are many parameters and options which have to be tuned by trial and error, such as the length of the insertion period and the coefficient $\lambda$. Moreover because of scale dependence of the adjustment process in the model there can be no universally applicable "best" values.

(b) It does not make best use of isolated or incomplete data (see 4d).

(c) It is not well suited for automatic quality control of data. Data checking must ultimately be performed using redundant information among the observations (comparison with the first-guess is insufficient since it is those data which disagree but are correct which contribute most). Since we are not usually so fortunate as to have much redundant information in the observations for one variable, checking is best done using a multivariate 3-dimensional method (see 4e). The complications this entails are not compatible
with the simplicity needed for repeated interpolation at reasonable cost. In practice quality control is also hindered by "noise" in the assimilation model fields, and the need to check all data before starting assimilation.

(d) Having to build the analysis round a forecast model is constraining. Models have natural tendency to expand to fill the computer resources available, reducing the scope for analysis techniques not readily compatible with the model's formulation. There is an incentive to design the analysis in terms of model variables (surface pressure, temperature, specific humidity), irrespective of whether these are best.

Neither these weaknesses nor the strengths of section 2 are unique to this method, but are attributes of all methods to a greater or lesser extent. Their relative importance will be dependent on the application, and the precise implementation.

4. Some examples (There is not room here to reproduce the accompanying figures.)

(a) Rainfall. Swinbank (1980) showed that rainfall patterns during the FGGE assimilations were in good agreement with observations, although perhaps not more so than GCM integrations. Global rainfall rates in forecasts from an assimilation IIIb analysis were similar to those during the assimilation, while those from the equivalent ECMWF analysis were initially lower; a similar result to that of Miyakoda et al (1978). However, forecast amounts from the NMC analysis were not initially low, and daily maxima decreased similarly during the forecasts from all 3 analyses.

(b) Large scale divergent wind. The time mean 200 mb velocity potential field, and the first 3 empirical orthogonal functions of the time variation during May - July 1979, agreed well for the ECMWF IIIb and the assimilation IIIa analyses. (ECMWF values were a little smaller because of adverse effects of their initialization on the first guess used for the analyses.) Other circulation statistics also agree.
Analysis and forecast comparison. Assimilation, ECMWF and MOC sea level pressure analyses for the N. Atlantic were very similar, for a case study using FGGE data for February, as were the vertical velocities at 700 mb (derived directly from the model's continuity equation for the assimilation, and from the non-linear normal mode initializations of ECMWF and MOC). 2-day forecasts from these were equally similar, none being visibly better.

Assimilation of wind data. The overall impact of aircraft wind data on an assimilation analysis and subsequent forecast, measured by the difference from a run without the data, was similar in nature and location to the impact of the same data on the ECMWF analysis scheme, but impacts on the forecasts were smaller in magnitude. An isolated observation in a jet stream over the Pacific caused an isotach maximum (greater than the observed wind) downstream of the datum, because of advection during the forward assimilation. Of this 30 m/s difference between analyses only 20 m/s was in the rotational wind field and none was geostrophically balanced by differences between the height analyses. Thus 6 hours later most of the impact had been carried away and dispersed, and only a 5 m/s difference, in approximate geostrophic balance with the height difference, remained. The ECMWF analysis of this case was closer to balance because of their multivariate height and wind analysis method, and a forecast from this gave smaller changes during the first 6 hours and a better forecast of the jet's position and strength after 2 days.

Quality control. The simple univariate analysis used for checking the data while making the FGGE IIIa analyses failed to detect a VTPR sounding at 180°W 46°S which was about 5° too warm. Repeated insertion of this, with few nearby observations, generated an entirely spurious surface low. This and the thickness pattern obtained by attempting to fit the VTPR soundings gave a 500 mb height field which was not geostrophically consistent with a nearby radiosonde wind. Thus a multivariate 3-dimensional quality control method would have been more likely to detect the bad observation.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Author(s)</th>
<th>Year</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>'A study of the dynamics of four-dimensional data assimilation. Tellus, 33, pp. 43-60.</td>
<td></td>
</tr>
</tbody>
</table>
ON THE COMPLETENESS OF MULTIVARIATE OPTIMUM INTERPOLATION

By Norman A. Phillips
National Meteorological Center

Since non-linear modal initialization (NLMI) is the last step before making a forecast, the preceding analysis step must anticipate this final change to the initial fields. If we use the theoretically-based Baer-Tribbia method of NLMI, in which fast modes are first set equal to zero, this means that only the slow modes in the analysis are meaningful. This, in turn, means that (1) the analysis must be of slow modes only, (2) that observations must be "corrected" by subtracting an estimate of their fast mode content, and (3) that statistical guidance of the analysis must be based on first-guess errors containing slow modes only.

If $Y_k$ denotes the vector of grid-point values, and $Z_p$ the observations, the multivariate optimum interpolation analysis can be written as $(k = 1, K)$:

$$Y_k(\text{an}) = Y_k(\text{fg}) + \sum_{p=1}^{P} [Z_p - Y_p(\text{fg})] \sum_{q=1}^{P} \overline{O_{pq}} Y_q,$$

where "fg" denotes first guess, $y_k$ is a grid-point first-guess error and $y_q$ is the first guess error for observation $Z_q$ (q also = 1, P). $O_{pq}$ is the inverse of the error covariance matrix that determines the OI analysis weights. It is assumed in (1) that all data is used for each analyzed variable $Y_k(\text{an})$.

Suppose we have a complete set of orthonormal functions $\phi_{\ell}(x), \ell = 1,K$. Let $s_\ell = \phi_\ell$ for $\ell = 1, L$ denote a subset of this and let $f_{\ell} = \phi_{\ell+L}$ for $\ell = 1, K-L$ denote the remaining members. Suppose that the first-guess

---

1A full version of this paper is scheduled for the October 1982 issue of the Monthly Weather Review.
error \( y_k \) is allowed to contain only \( s_k \) components, so that in any one realization we may write

\[
y_k = \sum_{\ell=1}^{L} \delta S \delta s_k (x_k) \tag{2}
\]

(\( \delta S \) denotes the error in the amplitude of component \( s_k \).) The covariance \( y_k y_q \) can then be expressed as

\[
y_k y_q = \sum_{\ell=1}^{L} s_k (x_k) \sum_{m=1}^{L} (\delta S \delta s)_m (x_q)\tag{3}
\]

If this is substituted for \( y_k y_q \) in the correction sums on the right side of (1), the orthogonality over \( k \) of any \( s_k (x_k) \) to any \( f_m (x_k) \) means that that correction sum now has no components in the subset \( f_k \). Under these circumstances we can apply any "filter" we wish to the OI analysis correction procedure by proper expression of the first-guess error covariance according to (3), and by using all data for each analyzed quantity. In the present context, of course, the desired filtering is obtained by identifying \( s_k \) and \( f_k \) with the "slow" and "fast" modes defined in NLMI.

This filtering theorem does not depend on the "data" \( Z_p \) for its validity. The accuracy of the analysis will however depend on \( Z_p \). In a previous paper (Phillips, 1982) I have shown how subtraction of fast mode estimates from data can be performed with acceptable accuracy even if \( Z_p \) is "good" enough to contain some information about the fast modes \( f_k \) that are filtered out at the beginning of the analysis and inserted only by the final NLMI step.

Reference

Ghil, et al. (1981) have recently brought to our attention again the existence of a well-developed technique of data assimilation called Kalman (or Kalman-Bucy) filtering (see also Petersen, 1968). It differs from the so-called "optimum interpolation" or "OI" data analysis procedure in use at several large NWP centers by its attempt to predict the detailed covariance structure of the first guess error fields. It does this by using linear prediction equations, which, in the meteorological context would have to be derived by linearizing the model forecast equations. This paper attempts to (a) present the K-B technique in a context simple enough to demonstrate its relation to operational practice and (b) to point out dangers in the K-B procedure that exist if the linear prediction equations are in error.

The following discretized linear two point system is used:

\[
\begin{align*}
T_{1n+1} & = [\nu T_1 - \mu T_2 + r_1]_n \\
T_{2n+1} & = [\mu T_1 + \nu T_2 + r_2]_n
\end{align*}
\]

where \( T_i \) are the true values. \( r_i \) is a noise that represents a deviation of the forecast model from the true behavior. If \( a_{in} \) and \( f_{in} \) represent analysis errors and forecast first guess errors, (1) leads to the prediction equations

\[ A \text{ more detailed version of this paper is given in Office Note 258, June 1982, of the National Meteorological Center W32, WWB, Washington, DC, 20233.} \]
The normal OI procedure produces the analysis errors

\[
\begin{align*}
\hat{a}_1 &= (1 - \alpha) \hat{f}_1 - \beta \hat{f}_2 \\
\hat{a}_2 &= (1 - \delta) \hat{f}_2 - \gamma \hat{f}_1 \\
\hat{a}_1 &= -\beta \hat{f}_2 + (1 - \alpha) \hat{f}_1 
\end{align*}
\]

in which the analysis weights \(\alpha, \beta, \gamma, \delta\) are readily derived functions of \(\hat{f}_1, \hat{f}_2\) and the observation errors \(\varepsilon_i \varepsilon_j\).

If \(r_1 r_2\) and \(\varepsilon_i \varepsilon_j\) are independent of \(n\) this system quickly approaches a steady state unless \(\sigma = \mu^2 + \nu^2 = 1\). Explicit results are readily derived when \(r_1 r_2 = r_2 r_2 = 0\), and \(\varepsilon_1 \varepsilon_1 = \varepsilon_2 \varepsilon_2 = 0\).

They show that for \(r^2 \to \infty\) or \(\varepsilon^2 \to 0\), the analysis error \(\hat{a}_1 a_1 = \hat{a}_2 a_2\) approaches \(\varepsilon^2 - \varepsilon^4 / r^2\). For the opposite case of \(r^2 \to 0\) and \(\varepsilon^2 \to \infty\), the analysis error approaches \(r^2 / (1 - \sigma)\) when \(\sigma < 1\) and \(\varepsilon^2 (1 - 1) / \sigma\) when \(\sigma > 1\). When \(\sigma = 1\) the asymptotic solution is singular, the analysis error approaches \(\varepsilon r\).

The system can be used to explore the usual practice of assuming a value for the correlation \(\rho = \hat{f}_1 \hat{f}_2 (\hat{f}_1 \hat{f}_1 - \hat{f}_2 \hat{f}_2)^{-1/2}\). If \(\rho\) is arrived at by "off-line" computation using true forecast errors, the results are almost as accurate as the true K-B results. However, the apparent values of
first guess error are a completely unreliable guide to obtaining an effective fixed value for \( \rho \). The system can also be used to explore the consequence of using incorrect values of the dynamic parameters \( \nu \) and \( \mu \). Underestimate of \( \sigma = \nu^2 + \mu^2 \) is particularly dangerous. Several stabilizing procedures to guard against this can be designed by adding extra "system noise" \( r^2 \). However, this can have the effect of reducing the detailed equations (2) to ones in which the dynamic matrix is almost replaced completely by added noise. The power of the K-B approach is therefore greatly weakened. (A more detailed study of this can be found in the book edited by A. Gelb.)

References


Normal mode initialization in the ANMRC data assimilation scheme

Kamal Puri

Australian Numerical Meteorology Research Centre,
P.O. Box 5089AA, Melbourne, Australia. 3001.

Using the ANMRC data assimilation system, Bourke et al. (1982) have found, as have other groups, that nonlinear normal mode initialization (NMI) leads to two undesirable effects namely,

1) The Hadley cell in the assimilation model is seriously depleted
2) There is considerable loss of surface pressure information

Two possible solutions to the Hadley cell problem have been found, one based on incremental linear NMI (Puri et al, 1982) and the second using nonlinear NMI (Puri and Bourke, 1982). The incremental linear scheme essentially subtracts out gravity modes from increments to the model state due to insertion of data. Thus for vertical mode n

\[ \Delta Y^I(n) = \Delta Y^A(n) - \sum_s \sum_j G \hat{Y}_j^s(n) Y_j^s(n) \]  \hspace{1cm} (1)

where \( \Delta Y^A(n) \) is the analysed increment to the model state, \( \Delta Y^I(n) \) is the linearly initialized increment, \( \hat{Y} \) is the Hough vector function, \( Y_j^s \) are the expansion coefficients, \( s \) is the zonal wave number and \( \sum_j^G \) denotes a summation over gravity modes only.

Thus for vertical mode n the initialized model state is

\[ Y^I(n) = Y^O(n) + \Delta Y^I(n) \]  \hspace{1cm} (2)

where \( Y^O(n) \) is the model state before the insertion of data. Since the incremental linear scheme does not directly affect
the background model fields it should preserve the meridional circulation built up by the model during data assimilation although it may not necessarily retain changes in the circulation implied by the data.

The nonlinear scheme referred to here as modified NMI is based on the finding that the Hadley cell in the nine level ANMRC spectral model is maintained by convective adjustment which mainly influences the low frequency gravity modes for vertical modes 3 and 4 and in particular vertical mode 4 (Puri and Bourke, 1982; Puri, 1982). Thus if only the first two vertical modes are initialized it is found that the Hadley cell is well retained. To initialize the higher vertical modes Puri and Bourke (1982) have introduced a frequency cutoff which excludes the low frequency modes responsible for maintaining the Hadley cell from the initialization.

Both incremental linear and modified nonlinear NMI schemes are found to retain the Hadley circulation in the model. This is shown in Fig. 1 which shows the latitude-height plots of the zonally averaged vertical velocity in an assimilation experiment after modified nonlinear NMI, standard nonlinear NMI and incremental linear NMI. The modified and standard nonlinear NMI schemes are found to be much more effective in controlling the gravity wave oscillations in the assimilating model than the incremental linear NMI. The level of noise in the linear scheme however is significantly less than that introduced by the insertion of data.

In an attempt to minimize the loss of surface pressure information during nonlinear NMI the variational formalism of constrained NMI developed by Daley (1978) is used. For each vertical mode n an attempt is made to minimize a variational integral subject to the constraint that the final model state satisfies Machenhauer's nonlinear NMI condition. Integrals of the form

$$I^n = \int [\omega_n^2 (\nabla \psi_n^2) + \omega_n^2 (\nabla \chi_n^2) + \omega_n^2 (\Delta p_n^2) ] dA$$
are minimized. Here $\psi$ and $\chi$ are the stream function and velocity potential respectively and $P = \Phi + RT_0(\sigma) \ln p^*$, $\omega_\psi$, $\omega_\chi$ and $\omega_p$ are pre-specified weights and $\Delta \psi^n$ etc. denote changes in the model fields during initialization. The choice of weights, which is crucial for constrained initialization, was based purely on the need to minimize the loss of surface pressure information. However, the mass weights must not be so high as to have a deleterious effect on the wind field; there clearly has to be a compromise between retaining as much mass information as possible and ensuring that the wind field is not badly damaged. The weights were allowed to vary with latitude only and north of $15^\circ S$ an unconstrained initialization was performed. There was also the option of varying the weights with vertical mode. Figure 2 shows a comparison of five day assimilation cycles using constrained and unconstrained nonlinear NMI. The main feature to note is that the constrained scheme leads to a reduction in the loss of surface pressure information without seriously affecting the wind field. Model forecasts from days 3 and 5 of the two cycles are found to be similar in terms of subjective and objective measures.

References


Fig. 1. Latitude-height distribution of zonally averaged vertical velocity scaled by factor 100 after modified nonlinear NMI (bottom), nonlinear NMI (middle) and incremental linear NMI (top). Units are mb s$^{-1}$, the dashed lines indicate negative values (upward flow) and the contour interval is 1 mb s$^{-1}$. The model sigma levels are indicated on the ordinate.
Fig. 2 The root mean square observation fitting errors for five day assimilation cycles using constrained and unconstrained initialization. The domain covered is from 25°S to 60°S.
Selected topics on the use of the variational methods in meteorology are presented. First four applications of variational methods are reviewed. These are: i) variational adjustment in discretization to conserve mass, total energy and potential enstrophy (Sasaki, 1976); ii) truncation error due to inconsistency of finite difference algorithm and can reach a significant magnitude especially in application of the variational methods (Sasaki, Ray, Goerss and Soliz, 1979); iii) use of inequality constraint in variational adjustment (Sasaki and McGinley, 1981); iv) computational stability of gradual application of constraint during the course of time integration (Sasaki, 1980). Discussion of the two new developments follows: i) the optimal determination of weights used in formulation of the weak constraint algorithm; ii) an efficient variational method using the distribution (hyperfunction) to solve the problem of pointwise observations in generating a continuous field.

Weight in Weak Constraint Formulation

Weights in functionals formulated with weak constraints may be determined in several different ways depending upon the nature of problem. In a spectral sense, the weight determines the low-pass filter that truncates high frequencies (Sasaki, 1969). In a statistical sense, these weights are inversely proportional to the square of the difference between analyzed and observed quantities (Stephens, 1965 and Sasaki, 1970).
In the latter case, Phillips (1981) showed a simple example of optimal
determination of weight when two variables X and Y are the solution of the
problem that is stated as;
\[
\min J : J = Q(X - X_0)^2 + P(Y - Y_0)^2
\]  
subject to the constraint (called "strong constraint"),
\[
Y = AX
\]
The quantity \(X_0\) and \(Y_0\) are prespecified observations for \(X\) and \(Y\) respectively. 
P and Q are the weight to be determined as well as \(X\) and \(Y\). Defining the analysis
errors \(\delta X\) and \(\delta Y\) as \(\delta X = X - X_t\) and \(\delta Y = Y - Y_t\),
where \(X_t\) and \(Y_t\) are the corresponding true values of \(X\) and \(Y\), Phillips showed that the weights \(P\) and \(Q\)
can be determined as
\[
P = \frac{1}{(\delta Y_o)^2} \quad \text{and} \quad Q = \frac{1}{(\delta X_o)^2}
\]  
where \(\overline{()\)}\) represents a sample mean of the vorticity \(\overline{()\)}\), and \(\delta X_o = X_o - X_t\) and
\(\delta Y_o = Y_o - Y_t\) (observation error). In the derivation of (3), Phillips assumed
there was no model error, i.e., \(Y_t = AX_t\), and no correlation between \(\delta X_o\) and \(\delta Y_o\).

The Phillips' theory can easily be applied to a case of weak constraint.
The most simple case is the one with one variable \(X\) and one weak constraint on
\(X\) as
\[
F(X) = AX + B = 0
\]  
Instead of (1) and (2) the problem may be stated as to find the solution \(X\) such
that
\[
\min J : J = Q(X - X_0)^2 + P(F(X))^2
\]
subject to (4). This corresponds to the case where \(Y = F(X) = 0\) and \(Y_o = 0\) in
the Phillips theory. Performing calculation of the first variation of the functional \(\{}\) in (5), we get
\[
Q(X - X_0) + P(AX + B)A = 0
\]
Then, solution $X$ is obtained as

$$ X = \frac{Qx_0 - PBA}{A^2p + Q} $$

(7)

The analysis error, $\delta X$ is

$$ \delta X = X - X_t = \frac{Q\delta x_0 - PA\delta x_m}{A^2p + Q} $$

(8)

where $\delta x_0 \equiv x_0 - x_t$ (observation error) and $\delta x_m \equiv Ax_t + B$ (Model error).

Assuming that there is no correlation between $\delta x_0$ and $\delta x_m$, i.e., $\overline{\delta x_0 \delta x_m} = 0$, the analysis error variance becomes

$$ \overline{\delta x^2} = (X - X_t)^2 = \frac{Q^2(\delta x_0)^2 + P^2(\delta x_m)^2}{(A^2p + Q)^2} $$

(9)

If the weights $P$ and $Q$ are chosen by setting condition that $\overline{\delta x^2}$ is to be minimized, then the optimal weights are given as

$$ P = \frac{1}{(\delta x_m)^2} \quad \text{and} \quad Q = \frac{1}{(\delta x_0)^2} $$

(10)

The weak constrain formulation has been used in a number of variational analyses of meteorological fields, initialization (some recent references are: Daley, 1978; Barker, 1982, Tribia, 1982) and assimilation (Sasaki and Goerss, 1982). The Phillips' theory and its extension may give some insight of the weights used in these studies.

A Variational Distribution Formulation with Pointwise Observations

Using the Dirac function, a pointwise observation $f_i$ at the location $(x_i, y_i)$ is expressed

$$ f_i = f(x_i, y_i) = \int f(x, y) \delta(x-x_i) \delta(y-y_i) \, dx \, dy $$

(11)
where \((x, y) \in \Omega, \delta(\xi)\) is the Dirac function that has properties; \(\delta(\xi) = 0\) where \(\xi \neq 0\) and \(\delta(0) = \infty\) where \(\xi = 0\), and it has a property that \(\int \delta(\xi) \, d\xi = 1\).

In 1946, Schwartz introduced the theory of distribution (also called hyperfunction, ideal function or generalized function). The distribution is a linear function \(f(\phi)\) defined for all \(\phi(x) = C_0(\Omega)\). Continuous function \(f(x)\) generates a distribution

\[
F(\phi) = \int_\Omega \phi(x)f(x) \, dx \quad (12)
\]

The Dirac function is a special form of the distribution when \(\phi(\xi) = \{\delta(\xi) : \delta(\xi) = \sum_{n=0}^{\infty} e^{i\pi n\xi/L}, \xi \in \Omega\}\). The expression \((\ )\) diverges, however it can be used following the calculus rules deduced by the theory of distribution. The theory is attractive to apply to meteorological problems because it deals with singularity and discontinuity and the calculus rules are very similar to those of ordinary functions. In meteorology, the distribution was first introduced to investigate anisotropic property of weight of the weak constraint formulation (Sasaki, 1971) by calculating the Green's function. It is further applied to study a pointwise observation problem. Consider a two-dimensional spline interpolation problem for simplicity;

\[
\min J; J = \int_{\Omega} \int_{\Omega} (\Delta f(x,y))^2 \, dx \, dy + \sum_{n=1}^{N} \lambda_i (f(x,y) - f_i) \quad (15)
\]

subject to

\[
f(x_i, y_i) = f_i, \quad i = 1, 2, ..., N \quad (14)
\]

where \(\Delta\) is the Laplacian operator, \(\Omega\) is chosen as a rectangular domain, a subspace of \(R^2, 0 < x < L_x, 0 < y < L_y\), and \(f_i\) is the given value of observation of the variable \(f\). Using the Lagrange multiplier theorem, (13) and (14) are written in a variational form that

\[
\min J; J = \int_{0}^{L_x} \int_{0}^{L_y} (\Delta f(x,y))^2 \, dx \, dy + \sum_{n=1}^{N} \lambda_i (f(x,y) - f_i) \quad (15)
\]
where $\delta$ is the first variation operator, $\lambda_i$ is the Lagrange multiplier.

For simplicity, we employ the boundary conditions that all of $f$ and its second derivatives vanish on the boundary (called simply supported edge boundary condition, Oden and Ripperger, 1981);

$$f(x,y) = 0 \quad \text{and} \quad \frac{\partial^2}{\partial x^2} f(x,y) = 0 \quad \text{on} \quad x = 0, L_x \quad (16)$$

$$f(x,y) = 0 \quad \text{and} \quad \frac{\partial^2}{\partial y^2} f(x,y) = 0 \quad \text{on} \quad x = 0, L_y \quad (16)$$

If the boundary conditions are applied to the above equation, all of the boundary terms are dropped. Then (15) becomes

$$\int_{0}^{L_x} \int_{0}^{L_y} \left( 2\Delta^2 f(x,y) + \sum_{i=1}^{N} \lambda_i \delta(x-x_i) \delta(y-y_i) \right) \delta f(x,y) \, dx \, dy = 0 \quad (17)$$

Using the sine harmonic series that satisfies the boundary conditions, $f(x,y)$ is expressed by

$$f(x,y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} F_{m,n} \sin \frac{m\pi x}{L_x} \sin \frac{n\pi y}{L_y} \quad (18)$$

where $F_{m,n}$ and $\lambda_i$ are;

$$F_{m,n} = \frac{1}{2\left( \frac{m\pi}{L_x} \right)^2 + \left( \frac{n\pi}{L_y} \right)^2} \lambda_i \sin \frac{m\pi x}{L_x} \sin \frac{n\pi y}{L_y} \quad (19)$$

$$\lambda_i = \int_{0}^{L_x} \int_{0}^{L_y} \left( \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{2\left( \frac{m\pi}{L_x} \right)^2 + \left( \frac{n\pi}{L_y} \right)^2} \sin^2 \frac{m\pi x}{L_x} \sin^2 \frac{n\pi y}{L_y} \right) \delta f(x,y) \, dx \, dy \quad (20)$$

Thus, particular solution was obtained (note that simplified boundary condition was considered). In order to satisfy conditions of a more general form, we need to add a homogeneous solution $f_h$ of biharmonic equation

$$\Delta \Delta f_h = 0 \quad (21)$$
with the boundary condition imposed on \( f(= f_p + f_h) \) and its derivatives. The Levy method was adapted to obtain the homogeneous solution (Oden and Ripperger, 1981).

Also, the "distribution method" was tested for the case of observation with bounded error and the case of weak constraint formulation.

Throughout these tests, the "distribution method" shows its superiority with respect to the computational accuracy and speed. The higher the order of derivatives used in smoothing, the lower the maximum wave number required in (18). By using \( 1 \leq m \leq 15 \) and \( 1 \leq n \leq 15 \), the tests gave very satisfactory results, containing the truncation error relative to the value itself being less than \( 10^{-4} \). The "distribution method" is very efficient computationally. A manuscript is in preparation to describe details of the "distribution method" and the results.

Acknowledgment

The work summarized in this extended abstract has been supported under the National Science Foundation Grant #ATM7824892 and the Navy Grant #N00014-79-C-0758.

References


FOUR-DIMENSIONAL ASSIMILATION AT GFDL USING FGGE DATA
by
William Stern

A continuous data assimilation system used for the processing of observations collected during GATE (Miyakoda et al., 1976), has been further developed and currently is being employed to produce analyses from FGGE data. The present system involves three main phases: data pre-processing, dynamic assimilation and initialization. A schematic overview of the system is shown in Fig. 1.

The first part of the pre-processing involves sorting the level II-b data by type, blocking it into 6-hour intervals and performing quality control checking. Next, insertion data is prepared by optimum interpolation (referred to as OPI hereafter) of data to 19 mandatory pressure levels and to the model grid points for u,v,T,q and also P_s. The OPI (Gandin, 1963) is univariate with a vertical range of 3 mandatory pressure levels (see Fig. 2) and a horizontal range of 250 km. Twelve hours of data are processed in 2 hour blocks via six separate OPI analyses. The first guess used for all OPI is the most recent synoptic time analysis available (00Z or 12Z). A maximum of 8 observations are used to determine each gridpoint OPI analysis value. Next, weights are assigned to the analysis values depending on the accuracy of the data involved. Finally, an interpolation in time is performed in order to fill in gaps at particular analysis points. By using earlier and later analysis values that are available, a smooth time history of insertion data is maintained.

The pre-processed insertion "data" is assimilated into the forecast model in a weighted manner (i.e., if no observations were used in the OPI at a particular gridpoint, the model solution would be retained
entirely, however, if the observations used were quite reliable the model solution would receive a low weight and should be essentially replaced by the insertion data.) The model being used here is a global spectral model rhomboidally truncated at 30 waves with 18 levels in the vertical, a Monin-Obukhov surface boundary layer formulation and diurnally varying solar radiation are included. For more details concerning the dynamics and physical processes see Gordon and Stern (1982). By the repeated insertion of the same data at each time step over a 2-hour interval, it is felt that a balanced state may be achieved that is faithful to the data, and is also consistent with the model's dynamics and physics. The model should impose its own constraints between the variables. For this reason a multi-variate OPI was not used. There is evidence that the model does accept some data, as the rms differences between the insertion values and the model solution are reduced after several assimilation time steps. However, there are also indications that some data are being projected into fast gravity modes and hence do not contribute to a balanced state.

The initialization has been designed to control the growth of spurious gravity modes but not to alter the model balance, especially in the tropics. As indicated in Fig. 1 by "N", a nonlinear normal mode initialization is performed every 6 hours. The numerical method used to balance the linear and nonlinear contributions to the tendencies is Machenhauer's (1977) with most of the details described by Bailish (1980). The GFDL scheme initializes the first 7 of 18 vertical modes and allows 4 iterations for convergence. In order not to risk initializing out gravity modes that might be important, especially in the tropics, only those modes with periods shorter than 6 hours are adjusted. This
scheme was motivated by noticing that most of the amplitude of surface pressure oscillations from un-initialized forecasts in the tropics seemed to be in periods longer than 6 hours. With the 6-hour cutoff imposed it turns out that most of the modes associated with vertical modes 6 and 7 are not being initialized and most of the modes for the lowest zonal waves of mode 5 are not initialized. If one looks at the vertical structure functions for the model (Fig. 2) it appears that the main contributions at the levels where a return Hadley circulation (or the quasi-global meridional flow) is found are modes 6 and 7 and to some extent 5. Since these modes are remaining essentially un-initialized for the long zonal waves, the Hadley circulation should not be affected by the initialization. This may be seen from Fig. 3. Here zonal mean meridional wind latitude-height profiles are compared for a 4D-analysis before any initialization, using the current scheme and extending the initialization cutoff frequency to 48 hours. It is quite clear that the 6-hour cutoff scheme maintains the strength of the return Hadley flow quite close to the un-initialized case, while the more conventional 48-hour cutoff suppresses the return circulation. There is an increase in the amplitude of the surface pressure oscillations with this selective initialization, but they are still reasonably controlled.

For most of the FGGE data processing the initialization appeared to control the amplitudes of spurious gravity modes and remain well behaved. However, there were two instances which resulted in a departure from the scheme described above. The first problem occurred at the start of the FGGE year, December 1, 1978. An NMC Level III analyses for December 1, 1978 00Z was used to begin the assimilation process. The top level at which NMC data is available is 50 mb, while the model requires data
at approximately 19 mb and 2.2 mb. The extrapolation, which does not impose any balance constraints, apparently excited some gravity modes that were not being initialized by the scheme with the 6-hour cutoff. Figure 4 shows latitude-height cross sections of zonal mean meridional velocity after two days of FGGE processing. The bottom panel shows a spurious stratospheric feature which appears immediately after starting up, with the selective initialization being applied every 6 hours. By initializing all gravity modes for the first 7 vertical modes at December 1, 1978 0Z and 12Z and then returning to the initialization with a 6-hour cutoff the amplitudes of the spurious stratospheric modes are greatly reduced as seen in the top panel of Fig. 4. It is interesting to note that despite the different start up initializations that the tropospheric features remain pretty much unaffected. In contrast to the first case where not enough modes were initialized the other episode is one in which "too many" modes were being initialized. During July GFDL's FGGE analyses exhibited a very strong S. Hemisphere polar night jet at about 65°S. Zonal mean speeds reached about 120 m sec\(^{-1}\) at the top level (this is probably unrealistically strong) and nearly 100 m sec\(^{-1}\) at level 2. Phillips (1981) demonstrated that under conditions where the advective time scale for a mode is on the same order as the period of the mode itself the Machenhauer iterative scheme will not converge. These conditions were met for several modes associated with the higher internal vertical modes (5 and 6), and the highest zonal wave No. (30). The effect was an amplification of the highest waves in several upper level analyses. The divergence of several modes may be seen in Fig. 5 where the mode tendency amplitude is plotted as a function of iterations for four individual modes, it is evident that their amplitudes are increasing with increasing iteration instead of converging
to a certain value. Phillips (1981) recommended the use of the Baer-
Tribbia initialization as a remedy. In practice for the current operational
FGGE processing system, instead of using the Baer-Tribbia method the
solution tentatively chosen was to only initialize the first three
vertical modes after the first iteration. This modified scheme has
remained well behaved.

Assessing the usefulness and quality of FGGE Level III-b analyses
will involve many studies by many groups throughout the world over the
next few years. At GFDL numerous daily extratropical height and wind
analyses have been monitored. In general they compare favorably with
the ECMWF analyses in the phase and amplitude of medium to long wave
features although there is some tendency for the ECMWF system to produce
somewhat deeper troughs in data sparse regions. The GFDL products
appear to exhibit more ageostrophic flow than those from the ECMWF, this
is probably related to the differences in the OPI analyses. As part of
a general circulation study monthly mean and variance quantities have
been computed for several FGGE months. They appear quite reasonable and
in some instances have been compared with similar quantities from the
ECMWF. Outside of the tropics both sets of circulation features agree
closely with slightly greater wind magnitudes at the jet level in the
GFDL analyses. There are also differences in the stratosphere above 50
mb but here both analysis are probably unreliable. In the tropics some
significant differences are apparent. The monthly mean 850 mb convergence
and 200 mb divergence over the Amazon Basin and Central Africa are much
stronger in GFDL's analysis, based on analyses from April 1979. This
may be attributed to differences in the convective parameterizations and
the initialization schemes. Also reflecting these differences in
divergent flow, is the latitude-height cross sections of meridional wind
The most striking difference is the much stronger mean return Hadley circulation in GFDL's products. This difference has also been noticed in other months.

The continuous data assimilation system appears quite viable. Its main virtues being the potential to achieve a consistency among variables that utilizes the full physics of the model and the ability to assimilate asynoptic data closer to their actual observation time. Some of the areas that could probably be improved upon are increased data acceptance and reduction of some noise in the analyses. Application of variational linear normal mode and improvements to the OPI are proposed to be investigated.

Acknowledgements

I am grateful to Dr. K. Miyakoda, Dr. N.C. Lau and J. Ploshay for their valuable input to this report. Thanks also go to Betty M. Williams for typing this paper.

References


Fig. 1: Schematic overview of the GFDL data processing system during FGGE.

Fig. 2: The first 7 vertical modes of the 18 level wave 30 spectral model using $T(\sigma)=250^\circ K$. 
Fig. 3: Zonal mean meridional wind for 4-D analyses from Dec. 7, 1978 00Z with no initialization (top), modes with periods ≤ 6 hours from the first 4 vertical modes initialized (middle), and modes with periods ≤ 48 hours from the first 4 vertical modes initialized (bottom). Contour interval is 0.5 m sec⁻¹.

Fig. 4: Zonal mean meridional wind for 4-D analyses from Dec. 3, 1978 00Z with all modes from the first 7 vertical modes initialized (top) and with 6 hour cutoff imposed on the initialization of modes from the first 7 vertical modes (bottom). Contour interval is 1 m sec⁻¹ and shaded regions are negative.
Fig. 5: Gravity mode tendency amplitudes vs. iteration for selected modes with vertical structures of modes 5 and 6.

Fig. 6: Monthly mean zonal mean meridional wind profiles for April 1979 based on GFDL analyses (top) and ECMWF analyses (bottom). Contour interval is .5 m sec⁻¹.
1. Introduction

Normal mode initialization has proved itself to be a highly successful technique for suppressing unwanted high-frequency noise generated during data assimilation; in the current jargon, the analyses are successfully projected onto the slow manifold (Leith, 1980). One slightly disturbing feature of this technique is that, like any other initialization scheme, it inevitably changes the analysed fields. If the changes are small relative to the expected analysis error, then this is of no consequence; but typically it is found that although the implied changes to the wind field are acceptably small, apart from the somewhat separate problem of the large-scale tropical divergence fields (Puri & Bourke, 1982), the changes to the mass field (principally the surface pressure) can be a little too large for comfort. This apparent imbalance is presumably due in part to deficiencies in the analysis scheme (e.g. inappropriate structure functions) and in the initialization procedure itself (e.g. absence of physics in the nonlinear forcing). Even if these deficiencies can be overcome, the problem is likely to remain to some extent, since there is a mismatch between geostrophic adjustment theory - which shows that for most scales the mass field will adjust to the rotational part of the wind field - and the fact that surface pressure is one of the most accurately observed and analysed variables.

Daley (1978) proposed a variational method to tackle this problem. The usual unconstrained initialization schemes alter the fast (gravity) modes in the analysis, leaving the slow (Rossby) modes untouched. Using a variational approach, we can also make compensating changes in the slow modes, in order to return closer (in some sense) to the original analysis, while remaining close
to the slow manifold. Since the evolution of the forecast is primarily governed by the slow modes (Daley, 1980), this technique may significantly change (and perhaps even improve) the forecast, unlike unconstrained initialization which merely removes the spurious high-frequency component.

2. Definition of the variational integral

In the case of a barotropic model, a suitable variational integral (Tribbia, 1982) is

$$ I = \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \left\{ w_v |\Delta \vec{v}|^2 + w_h (\Delta h)^2 \right\} d\lambda \cos \theta d\theta $$

where $\Delta \vec{v}$ and $\Delta h$ are respectively the changes made by the initialization to the wind field and the height field, the latter being scaled so that $(|\vec{v}|^2 + h^2)$ is proportional to the total energy. The weights $w_v$ and $w_h$ reflect our confidence in the wind and height analyses; ideally they should be functions of $\lambda$ and $\theta$, but to keep the problem computationally tractable, here they will be functions of $\theta$ only.

For a multi-level model, let us assume that the vertical normal modes are orthogonal. In the case of the ECMWF gridpoint model (Temperton & Williamson, 1981) this is true only if the vertical modes are derived using an isothermal base atmosphere. If we minimize a variational integral

$$ I(m) = \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \left\{ w_v |\Delta \vec{v}|^2 + w_h (\Delta h)^2 \right\} d\lambda \cos \theta d\theta $$

for each vertical mode $m$, where $\vec{v}$ and $h$ are coefficients of the vertical normal modes, then using the energy relationships derived by Kasahara and Puri (1981) it can be shown that the three-dimensional variational integral being minimized is

$$ I = \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} \int_0^{2\pi} \left\{ w_v (\Delta \vec{v})^2 + w_h \left[ \frac{c_p}{T} (\Delta T)^2 + \frac{T}{c_p} (\Delta \rho \vec{v})^2 \right] \right\} d\lambda \cos \theta d\theta d\sigma $$

where $T$ is the temperature of the isothermal base atmosphere.
3. **Computational procedure**

For each vertical mode $m$, we minimize a variational integral $I(m)$ given by Eq (1). Using the orthogonality of the Fourier modes in longitude, this is equivalent to minimizing

$$I(m, k) = \int_{-\pi/2}^{\pi/2} \left\{ w_v |\Delta \hat{\gamma} |^2 + w_h (\Delta \hat{h})^2 \right\} \cos \theta \, d\theta$$

for each wavenumber $k$, where $\hat{\gamma}$ and $\hat{h}$ are Fourier coefficients.

Discretizing Eq (2) and separating symmetric and antisymmetric modes (again using orthogonality), we find that

$$I(m, k) = || W^{1/2} \Delta \hat{\gamma} ||_2$$

where $W$ is a diagonal matrix containing the weights, and $\hat{\gamma}$ is a vector containing values of $\hat{u}$, $\hat{v}$, $\hat{h}$ running from pole to equator as found in the derivation of the horizontal normal modes - see for example Section 2b of Temperton & Williamson (1981).

The standard initialization procedure specifies a change in $\hat{\gamma}$ due to changes in the gravity mode coefficients, $(\Delta \hat{\gamma})_G$. In order to minimize $I(m, k)$, we define a change due to the Rossby modes,

$$(\Delta \hat{\gamma})_R = E_R \Delta \xi_R$$

where $E_R$ is a matrix whose columns are the Rossby mode eigenvectors, and $\Delta \xi_R$ is a vector of changes to the Rossby mode coefficients, to be determined.

The problem of minimizing $I(m, k)$ then reduces to solving the linear least-squares problem

$$W^{1/2} E_R \Delta \xi_R = - W^{1/2} (\Delta \hat{\gamma})_G$$

The procedure can be formulated so that no extra work is required in comparison with unconstrained normal mode initialization, although some additional matrices have to be stored.
This variational initialization scheme can be iterated, in which case it is convenient to define the variational integral $I$ in terms of the changes made during each iteration, as in Daley (1978).

4. **Summary of results**

An extreme case of variational normal mode initialization is obtained by setting $w_v = 1$, $w_h = 0$, i.e. forcing the mass field to fit the analysed wind field. The procedure then converges well, though as expected it produces rather larger changes in the mass field than the unconstrained scheme (which can be shown to be equivalent to setting $w_v = w_h = \frac{1}{2}$).

If on the other hand we set $w_v = 0$, $w_h = 1$, i.e. force the wind field to fit the analysed mass field, then catastrophic changes are made to the wind field in the tropics, and the procedure soon diverges, reflecting the violation of an ellipticity criterion (Tribbia, 1981).

Mixed schemes which weight the wind field more heavily in the tropics and the mass field more heavily in middle and high latitudes converge well and, outside the tropics, reduce the changes to the mass field without significantly damaging the wind field. Forecasts run from analyses initialized in this way are just as noise-free as those run from conventionally initialized states.

It would be highly desirable to let $w_v$ and $w_h$ vary as functions of $\lambda$, $\Theta$ and $\sigma$, but the problem then becomes computationally much more difficult. Tribbia (1982) has successfully treated the two-dimensional problem for a low-resolution barotropic model, but the full three-dimensional problem remains to be tackled.
References


ON THE EXISTENCE OF SOLUTIONS TO MACHENHAUER'S NON-LINEAR NORMAL MODE INITIALIZATION

Lennart Thaning

In this study the existence of solutions to the non-linear normal mode initialization proposed by Machenhauer (1977) and some of their properties, are examined in a low order model. The model is constructed by selecting three normal modes from a shallow water model on an equatorial $\beta$-plane (Matsuno, 1966, Tribbia, 1979). The selected modes are: one Rossby mode with zonal wavenumber $m = k$, one eastward propagating gravity mode with zonal wavenumber $m = 2k$ and one sloshing mode (which oscillates with a typical gravity frequency) with zonal wavenumber $m = 0$. All three modes are symmetric around the equator. The choice of normal modes is dictated by the following reasons:

a) The model shall be able to describe a balance between the wind and height fields that resembles the real atmosphere

b) The model shall contain two different time scales, one slow and one fast

c) The model shall contain non-trivial non-linear interactions between the motions on the different time scales
d) The non-linear normal mode initialization must generate non-trivial gravity fields.

We introduce the following notations: $X_r + iX_i$ for the complex sloshing amplitude; $Y_r + iY_i$ for the Rossby amplitude; $Z_r + iZ_i$ for the gravity amplitude. The model, without non-adiabatic forcing, can then be written (cf Thaning, 1982):

$$\frac{d}{dt} X_r = \Upsilon_1 X_i - \kappa_2 (Y_r^2 + Y_i^2) - \kappa_3 (Z_r^2 + Z_i^2)$$  \hspace{1cm} (1)

$$\frac{d}{dt} X_i = -\Upsilon_1 X_r - \kappa_1 X_r X_i$$  \hspace{1cm} (2)

$$\frac{d}{dt} Y_r = \Upsilon_2 Y_i - \beta_1 X_r Y_r + \beta_2 X_i Y_i - \beta_3 (Y_r Z_r + Y_i Z_i)$$  \hspace{1cm} (3)

$$\frac{d}{dt} Y_i = -\Upsilon_2 Y_r - \beta_1 X_r Y_i + \beta_2 X_i Y_r - \beta_3 (Y_r Z_i - Y_i Z_r)$$  \hspace{1cm} (4)

$$\frac{d}{dt} Z_r = \Upsilon_3 Z_i - \zeta_1 X_r Z_r - \zeta_2 (Z_r^2 + Z_i^2) + \zeta_3 X_i Z_i$$  \hspace{1cm} (5)

$$\frac{d}{dt} Z_i = -\Upsilon_3 Z_r - \zeta_1 X_r Z_i - 2 \zeta_2 Y_r Y_i - \zeta_3 X_i Z_r$$  \hspace{1cm} (6)

In these equations the $\Upsilon$:s, $\beta$:s and $\zeta$:s are the interaction coefficients describing the non-linear wave-wave interactions and the $\Upsilon$:s are the linear frequencies of the different modes.

In this model the initial state according to Machenhauer's unconstrained non-linear normal mode initialization is defined by:

$$\frac{d}{dt} X_r = \frac{d}{dt} X_i = \frac{d}{dt} Z_r = \frac{d}{dt} Z_i = 0$$  \hspace{1cm} (7)

By choosing $Y_i = 0$ one can show (provided that the signs of some of the interaction coefficients are known) that the initial state according to Equation (7) is defined by:
In a $X_i/Z_i$-plane, Equations (9) and (10) represent a parabola and a hyperbola and from Figure 1, (in which the signs, but not necessarily the magnitudes, of the frequencies and the interaction coefficients are correct) we see that generally there are three possible initial states that satisfy the initial condition.

The characteristics of these possible initial states, for reasonable values of $Y_r$, are:

1) For the solution with small gravity amplitudes
   a) $\nabla_s \kappa_i \approx \kappa_3 u_r^2$; $\nabla_s z_i \approx \kappa_3 u_r^2$
   b) $\kappa_i \ll u_r$; $z_i \ll u_r$
   c) When integrated in time, the model stays close to this solution.

2) For the solutions with relatively large gravity amplitudes
   a) $\nabla_s \kappa_i \approx \kappa_3 z_i^2$; $\nabla_s z_i \approx -\kappa_3 \kappa_i z_i$
   b) $\kappa_i \gg u_r$; $z_i \gg u_r$
   c) When integrated in time, the model immediately leaves these states and large high frequency oscillations arise.

It can furthermore be shown that the iterative procedure normally used to solve the non-linear system of equations defining the initial state, can converge only to the state denoted by I in Figure 1.
It can be seen from Equations (9) and (10) that when $Y_r$ is increased, the parabola will be lifted and the hyperbola will be shifted away from its asymptote as well as from the $X_i$-axis. This means that the two solutions characterized by $Z_i < 0$ will become closer and for a certain critical value of $Y_r = Y_{r\text{crit}}$, they will cease to exist. $Y_{r\text{crit}}$ decreases with decreasing depth of the fluid but still for a depth of 0.9 m, it is not small enough to make it likely that the divergence of the non-linear normal mode initialization procedure for small values of the equivalent depth (cf. Temperton and Williamson, 1979), is caused by the "Rossby forcing" of the gravity modes. Since, external forcing appears in the gravity equations in the same way as the "Rossby forcing" a critical external forcing, beyond which only one initial state exists, can be calculated from the values of $Y_{r\text{crit}}$. As an example, Table 1 shows the critical values of the forcing on the v-momentum in wavenumber 0, at $30^0$ latitude ($k =6$).

<table>
<thead>
<tr>
<th>Depth H (m)</th>
<th>Critical forcing (m/s/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8792.8</td>
<td>-147.0</td>
</tr>
<tr>
<td>879.3</td>
<td>- 25.7</td>
</tr>
<tr>
<td>87.9</td>
<td>5.5</td>
</tr>
<tr>
<td>8.8</td>
<td>1.1</td>
</tr>
<tr>
<td>0.9</td>
<td>0.2</td>
</tr>
</tbody>
</table>

In a baroclinic model the external forcing above may be thought of as the vertical coupling between different vertical modes or non-adiabatic forcing. Therefore the figures in Table 1 indicate, that the non-linear normal mode initialization procedure diverges for small values of the equivalent depth, because the meteorologically relevant solution does not exist.
REFERENCES


Figure 1. The Machenhauer condition.

The parabola illustrates $\frac{dx_r}{dt} = 0$

The hyperbola illustrates $\frac{dz_r}{dt} = 0$

The intersections represent states that satisfy the Machenhauer condition.
I. Introduction

The extension of normal mode initialization through the addition of variational techniques is discussed. This problem has been considered by a number of others, most recently including Daley (1978), Phillips (1981), Temperton (1981), and Tribbia (1982).

Before examining the experiments performed using variational normal mode initialization (VNMI), it is useful to consider the motivation for considering this embellishment of standard normal mode techniques. A non-inclusive list of reasons is: 1) VNMI is a (partial) cure for the alteration of presumably accurate data over areas of dense data coverage which can occur using standard unconstrained NMI; 2) VNMI can be used to offset the tendency towards geostrophic adjustment of unbalanced observations so that the initialized fields better reflect the relative accuracy of wind and height measurements; 3) VNMI can be used to obtain balanced supplemental information, e.g., obtaining height analyses from the knowledge of the wind field alone; 4) Lastly, and probably most importantly, VNMI affects the initial state in the temporally coherent portion of the flow field (VNMI alters the slow manifold projection of the initial field from the projection obtained from unconstrained NMI).

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.
II. The Variational Problem

The mathematical problem of obtaining VNMI may be stated as follows: defining the vector of rotational (gravitational) mode coefficients as $\mathbf{x}(y)$, VNMI minimizes the fidelity metric

$$I = \int [(U_I - U_0)^2 + (V_I - V_0)^2] w_U + (\phi_I - \phi_0)^2 w_\phi \, ds$$

in which $U_I, V_I, \phi_I (U_0, V_0, \phi_0)$ are the initialized (observed) zonal, meridional wind components and geopotential, and $w_U$ and $w_\phi$ are weights reflecting the confidence in the observations of the winds and heights, respectively, subject to the constraint that initialized fields satisfy the normal mode balance condition $\mathbf{y} = \mathbf{G}(\mathbf{x}, \mathbf{y})$ depicted figuratively in diagram 1. The above constraint is nonlinear and thus the variational problem is also nonlinear and so approximate schemes for the solution must be utilized.

III. Low-Order Model Assessment

The nature of the various approximations used by the authors cited in the Introduction may be assessed to some extent by using a low-order model developed for initialization studies by Tribbia (1981) and subsequently used in a variational context by Phillips (1981).

For this model with three spectral variables $(U, V, \phi)$, the fidelity metric can be taken as

$$I = [(U_I - U_0)^2 + (V_I - V_0)^2] w_U + (\phi_I - \phi_0)^2 w_\phi$$

and the constraint is given by

$$V_I = 0 \quad U_I + BU_I^2 + \lambda \phi_I = 0$$
(κ, B are model constants reflecting the length scale relative to the deformation radius and the Rossby number respectively.) The various approaches to the solutions of the variational problem can be shown to correspond to the following polynomial equations in \( U_I \)

**Exact:**

\[
W U_i + \left( \frac{W\phi}{\lambda^2} \right) (1 + 2BU_i)(1 + BU_i)U_i = \begin{array}{l}
 W_U U_i - W_\phi \left( \frac{1 + 2BU_i}{\lambda} \right) \phi_0; \\
 \phi_i = -1/\lambda(U_i + BU_i^2)
\end{array}
\]

**Geostrophic:**

\[
W U_i + \left( \frac{W\phi}{\lambda^2} \right) U_i = W_U U_i - \frac{W_\phi \phi_0}{\lambda}; \\
\phi_i = -1/\lambda(U_i + BU_i^2)
\]

**Daley:**

\[
W U_i + \left( \frac{W\phi}{\lambda^2} \right) (1 + BU_i) = W_U U_i - \frac{W_\phi \phi_0}{\lambda}, \\
\phi_i = -1/\lambda(U_i + BU_i^2)
\]

**Phillips:**

\[
W U_i (1 + \frac{B}{\mu^2} U_i) + W_\phi (1 + \frac{B}{\mu^2} U_i) U_i = W_U U_i - \frac{W_\phi \phi}{\lambda}
\]

where \( \mu^2 = 1 + \lambda^2 \), \( \hat{U} = U_o - U_A \), \( \phi = \phi_o - \phi_A \), with \( BU_A^2 + (\mu^2 + 2BU_R)U_A + BU_R^2 = 0 \), \( \phi_A = \lambda U_A \), and \( U_R = (\lambda^2 U_o - \lambda \phi_o)/\mu^2 \).

A typical example of the solutions to these equations is shown in Table 1. This is representative of an ensemble of cases tested, and the results may be summarized by noting that the Daley, Phillips and exact solutions are essentially identical, with the Phillips technique giving a slightly more accurate minimization than the Daley approximation. The geostrophic approximation is least accurate due to its use of the erroneous geostrophic constraint.

**IV. Real data experiments with a coarse resolution shallow-water model**

The experiments reported on in this and the following section use a R20 spectral Hough function global shallow-water model with a mean free
surface height, \( H = 1 \text{ km} \), for forecasting and initialization. The observed data are taken to be the NMC FGGE 3a 500 mb \( u, v, z \) analysis for 0000 GMT December 13 1979. These fields are first initialized using a single iteration of unconstrained NMI and the post processed by adding 20 m and 5 m/s r.m.s. random perturbations to the height and wind fields respectively to simulate the combined effects of instrumental and analysis error.

Two sets of experiments were performed; the first used globally homogeneous weights, i.e. \( w_U \) and \( w_\phi \) invariant with respect to location with globe. The globally homogeneous experiments compared the results of unconstrained NMI with height constrained \( (w_U \sim 0) \), wind constrained \( (w_\phi = 0) \) and observationally constrained \( (w_U \text{ and } w_\phi \text{ prescribed from the known statistics of the error distribution}) \). The RMS and max differences between these initializations are shown in Table 2. Spectral information was also calculated and the results may be capitated as follows: The differences between observationally constrained and unconstrained initialization had a flat wavenumber spectrum while wind constrained and height constrained had energy difference spectral shapes which increased and decreased respectively with wavenumbers above the radius of deformation wavenumber \( \sim m = 6 \). All of the above is consistent with geostrophic adjustment considerations.

The second experiments used longitudinally varying profiles of \( w_U \) and \( w_\phi \) to model the effect of greater data accuracy over continental regions as opposed to oceanic regions \( (w_U \text{ and } w_\phi \text{ were taken to be } 10^3 \text{ times their oceanic values in regions of high confidence}) \). In this case, the homogeneous observationally constrained initialization was compared to the longitudinally varying constrained initialization and the difference
between these fields was contained entirely over the "oceanic" domain. A further experiment was performed solving the variational problem only over the long wave portion of the spectrum and using unconstrained WMI for the shorter scales \((m > 15)\). The RMS and max difference comparison is shown in Table 3. Spectral information was obtained showing that both longitudinally constrained VNMI's differed from the homogeneously constrained NMI in the largest scales. Furthermore, while the complete and spectrally restricted longitudinally varying constrained VNMI's differed from each other, the differences were smallest in the long waves.

V. Real data experiments - Forecast results

The various above-mentioned VNMI's were next used as initial conditions in short-range (3 day) forecast experiments to ascertain the effect of the initialization differences on forecast evolution. As a base result, a comparison between the uninitialized and the unconstrained NMI fields was made to demonstrate the effect of unconstrained initialization on a forecast. This comparison is shown in Table 5. An examination of the spectral energy difference shows that for the forecast period differences remain almost entirely in the gravitational modes.

Next, forecasts from the uninitialized homogeneous observationally constrained and longitudinally constrained VNMI's were compared to forecasts with unconstrained normal mode initial conditions. The RMS and max results from this comparison are shown in Table 6. The spectral evolution showed some tendency for small scale \((m > 14)\) differences energy to amplify while the larger scale differences tended to remain constant in amplitude over the integration period in agreement with predictability expectations.
VI. The above investigation has barely scratched the surface as to the experiments which may and should be attempted with VNMI. Among other interesting paths that might be pursued are the effect of initial "guess" on the variational iteration and improved mathematical techniques for the solution of the variational problem including the use of Monte Carlo techniques on the slow manifold, the possibility of combining local variational or OI structure analysis with a final variational initialization and, lastly, the use of wavenumber dependent confidence weights leading to integral kernel weights in the variational metric.

REFERENCES


Fig. 1 Geometric interpretation of variational normal mode initialization. Ellipses depict constant values of the fidelity metric. After Leith (1980).
### Table 1

<table>
<thead>
<tr>
<th></th>
<th>$U_I$</th>
<th>$\phi_I$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact</td>
<td>61.0</td>
<td>-58.2</td>
<td>1.64</td>
</tr>
<tr>
<td>Geostrophic</td>
<td>70.6</td>
<td>-55.9</td>
<td>12.31</td>
</tr>
<tr>
<td>Daley</td>
<td>60.7</td>
<td>-57.9</td>
<td>1.67</td>
</tr>
<tr>
<td>Phillips</td>
<td>61.1</td>
<td>-58.3</td>
<td>1.64</td>
</tr>
</tbody>
</table>

$U_O = U_B + \delta U = 60 \text{ m/s}; \phi_O = \phi_B + \delta \phi = -5710 \text{ m}^2/\text{s}^2; \delta U = 5 \text{ m/s}, \delta \phi = -200 \text{ m}^2/\text{s}^2. H = 1 \text{ km}, (\omega_z)^{-1/2} = 5 \text{ m/s}, (\omega_z)^{-1/2} = 200 \text{ m}^2/\text{s}^2$, and all velocities in m/s and all heights in meters.

### Table 2

Global homogeneous wts

<table>
<thead>
<tr>
<th></th>
<th>MAX</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z-U$</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>$V-U$</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>$O-U$</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

### Table 3

Longitudinally varying wts

<table>
<thead>
<tr>
<th></th>
<th>MAX</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L-0$</td>
<td>57</td>
<td>5</td>
</tr>
<tr>
<td>$L-LR$</td>
<td>27</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 4

Forecast - Initialized vs. uninitialized

<table>
<thead>
<tr>
<th></th>
<th>Day = 0</th>
<th>Day = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R-U$</td>
<td>$U$</td>
<td>$V$</td>
</tr>
<tr>
<td>$R-V$</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

### Table 5

Forecast differences

<table>
<thead>
<tr>
<th></th>
<th>Day = 0</th>
<th>Day = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R-U$</td>
<td>$U$</td>
<td>$V$</td>
</tr>
<tr>
<td>$R-V$</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>$L-U$</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>$O-U$</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

In this work, we exploit a duality between optimum interpolation and variational objective analysis to set up a certain variational approach to objective analysis, as well as to propose an approach to combining objective analysis and variational initialization into a single step.

In Section 2 we describe the duality. In Section 3 we describe the variational approach to objective analysis in the context of the estimation of vorticity and divergence from observed wind vectors. The duality of Section 2 is used to suggest how data concerning the energy spectral distribution in the atmosphere, as studied by, e.g., Baer (1974, 1981), Kasahara (1976), Kasahara and Puri (1981), Stanford (1979), may be used to help choose the form of the variational problem to be solved. The variational problem we solve to estimate vorticity and divergence has one "bandwidth" parameter (related to the half power point of the equivalent low pass filter) and one "partitioning" parameter, representing the relative allocation of energy to the divergent and non-divergent part of the wind. Both can be estimated by generalized cross validation (GCV).

In Section 4 we begin a synthesis of several ideas - i) the duality of Section 2, ii) the idea of partitioning the "signal" into divergent and non-divergent parts (which generalizes to the idea of partitioning the signal into slow and fast modes) and iii) the modified Kalman filter ideas as proposed by Ghil and coworkers (1981). The result is an argument that (in principle) one can set up the problem of estimating initial conditions by combining i) the forecast, ii) the observational data, iii) possibly certain physical constraints and iv) (partial) prior information concerning atmospheric spectral energy distribution, into a single variational problem. This result is in some sense a converse of Phillips (1982b), who argues that normal mode initialization can be done as part of optimum interpolation. The resulting variational problem as we propose it will have (at least) one bandwidth parameter, one balance parameter controlling the relative weight to be given to forecast data and observational data, and one partitioning parameter, governing the relative energy in the "signal" assigned to fast and slow modes. We conjecture that these three (control) parameters can be estimated dynamically from the data by GCV. They can also be chosen by trial and error. Objective analysis and (linear) normal mode initialization are thereby combined in one step. The estimation of one (or more) balance parameter(s) from the data may possibly avoid the pitfalls due to misspecification of the Kalman filter statistics as noted by Phillips (1982a).

Some of this work is joint with D. R. Johnson.
2. ON A DUALITY BETWEEN OPTIMUM INTERPOLATION AND VARIATIONAL OBJECTIVE ANALYSIS

We will describe this duality for the analysis of a univariate variable on the sphere (say 500 mb height) although the result is completely general. Let P denote a point on the sphere and let h(P) be (say), the 500 mb height minus the global average 500 mb height at P. Let the observations \( y_1, \ldots, y_n \) be modelled as

\[
y_i = h(P_i) + \epsilon_i,
\]

(2.1)

where the \( \epsilon_i \) are supposed to be zero mean independent measurement errors with common variance \( \sigma^2 = \text{E}\epsilon_i^2 \).

We suppose \( \text{E} h(P) = 0 \) and \( h \) has a prior covariance \( bR(P,Q) \),

\[
\text{E} h(P)h(Q) = bR(P,Q).
\]

Then using standard results in multivariate analysis

\[
\text{E}(h(P)|y_1, \ldots, y_n) = (R(P,P_1) \ldots R(P,P_n))(R_n + n\lambda I)^{-1} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},
\]

(2.2)

where \( \lambda = \sigma^2/nb \) and \( R_n \) is the \( n \times n \) matrix with \( ij \) th entry \( R(P_i,P_j) \).

Considering the expression on the right of (2.2) as a function of \( P \), which we denote by \( h_\lambda(P) \), it is easy to see that if \( \lambda = 0 \), then \( h_0(P) \) interpolates to the data exactly, \( h_0(P_i) = y_i, i = 1,2,\ldots,n \), whereas for \( \lambda > 0 \), \( h_\lambda(P) \) smooths the data, and \( \lambda \) controls the amount of smoothing - \( \lambda \) is the "bandwidth" parameter, \( h_\lambda(P) \) evaluated at grid points is the "optimum interpolant" of Gandin, given that all the available data points are used simultaneously.


For every covariance \( R(P,Q) \) satisfying \( \int R^2(P,Q)dPdQ < \infty \), there is a variational problem for which \( h_\lambda(P) \) is the solution. It is: find \( h \) in HR (a certain reproducing kernel Hilbert space) to minimize

\[
\frac{1}{n} \sum_{i=1}^{n} (y_i - h(P_i))^2 + \lambda J(h).
\]

(2.3)

where \( J(h) \) is the square norm of \( h \) in HR.

We will now describe \( J(h) \) in a manner which we hope will make its meteorological usefulness clear. \( R(P,Q) \) being a covariance, is a symmetric non-negative definite function and (given that it satisfies the hypothesis) has a so called Mercer-Hilbert-Schmidt expansion (Riesz -Sz.-Nagy (1955)) in its eigenfunctions and eigenvalues.
\[ R(P,Q) = \sum_{\lambda, s} \lambda_{\lambda} Y_{\lambda} S(P) Y_{\lambda} S(Q) , \text{ where } \int R(P,Q) Y_{\lambda} S(Q) dQ = \lambda_{\lambda} Y_{\lambda} S(P). \] (2.4)

(This expansion is a generalization of the factorization of a covariance matrix \( \Sigma \) as \( \Sigma = \Gamma D \Gamma^t \) where \( \Gamma^t = I \) and \( D \) is diagonal). We have deliberately used a notation to suggest that the eigenfunctions of \( R \) are spherical harmonics, but that is by no means necessary. In any case

\[ J(h) = \sum_{\lambda} \frac{\lambda_{\lambda}^2}{\lambda_{\lambda}}, \text{ where } \lambda_{\lambda} = \int h(P) Y_{\lambda} S(P) dP. \] (2.5)

(If \( h \) were a vector and \( R \) a matrix, then we would have \( J(h) = h^t R^{-1} h \)). As an example, if the \( Y_{\lambda} S \) are the spherical harmonics and \( \lambda_{\lambda} = (\lambda(\lambda+1))^{-2m} \) then by using the fact that the spherical harmonics are the eigenfunctions of the Laplacian, \( \Delta Y_{\lambda} S = -\lambda(\lambda+1) Y_{\lambda} S \), it is not hard to show that \( J(h) = \int |(\Delta^m h)^2| dP. \)

More generally, if \( \lambda_{\lambda} = \sum_{\nu=0}^{m} \alpha_{\nu} [\lambda(\lambda+1)]^\nu - 2 \), then

\[ J(h) = \int \sum_{\nu=0}^{m} \alpha_{\nu} (-\Delta)^\nu h |^2 dP. \] (2.6)

More details may be found in Wahba (1981a, 1981b), the use of Hough functions instead of spherical harmonics is briefly described in Wahba (1981b). If

\[ E h(P) h(Q) = b R(P,Q), \] (2.7)

then \( h \) has a Karhunen-Loève expansion

\[ h(P) = b \sum_{\lambda, s} h_{\lambda} S Y_{\lambda} S(P) \text{ where } bh_{\lambda} S = \int h(P) Y_{\lambda} S(P) dP, \] (2.8)

and the \( h_{\lambda} S \) are independent zero mean random variables with \( E h_{\lambda} S h_{\lambda} S' = b\lambda_{\lambda} S, \lambda S = \lambda S' = 0 \) otherwise. The \( \{\lambda_{\lambda} S\} \), that is, the relative energy distribution by wave numbers may be estimated from historical data and/or obtained (roughly) from theory, see Baer (1974, 1981), Kasahara (1976), Kasahara and Puri (1981), Stanford (1979).

3. ESTIMATION OF DIVERGENCE AND VORTICITY FROM THE OBSERVED WIND FIELD USING VECTOR SPLINES ON THE SPHERE. PARTITIONING OF THE DATA INTO DIVERGENT AND NON-DIVERGENT PARTS

Given observed wind data \((u_i, v_i)\) at point \( P_i, i = 1, 2, \ldots, n \), we estimate the vorticity and divergence as follows. The stream function and velocity potential are expanded in spherical harmonics

\[ \psi(P) = \sum_{\lambda=1}^{L} \sum_{s=-\lambda}^{\lambda} a_{\lambda} S Y_{\lambda} S(P), \quad \phi(P) = \sum_{\lambda=1}^{L} \sum_{s=-\lambda}^{\lambda} b_{\lambda} S Y_{\lambda} S(P). \] (3.1)
Then (for given \( \delta, \lambda \)), we find \( \{a_{\lambda S}, b_{\lambda S}\} \) to minimize

\[
\sum_{i=1}^{n} \left( \frac{1}{a} \frac{\partial \psi}{\partial \phi} \frac{1}{a \cos \phi} \frac{\partial \phi}{\partial \lambda} + \frac{1}{a} \frac{\partial \phi}{\partial \lambda} \right) - \frac{1}{(P_i) - u_i} - \frac{1}{(P_i) - v_i} + \frac{1}{\alpha a} \frac{\partial \phi}{\partial \lambda} - \frac{1}{\alpha a} \frac{\partial \phi}{\partial \lambda}
\]

\[ \sum_{i=1}^{n} \left( \frac{1}{a} \frac{\partial \psi}{\partial \phi} \frac{1}{a \cos \phi} \frac{\partial \phi}{\partial \lambda} - \frac{1}{(P_i) - u_i} - \frac{1}{(P_i) - v_i} \right)^2 + \frac{1}{\lambda} \left[ J_1(\psi) + \frac{1}{\delta} J_2(\phi) \right] \]

where

\[
J_1(\psi) = \sum_{\lambda=1}^{1} a_{\lambda S}^2 / \lambda_{\lambda S}^{(1)}, \quad J_2(\phi) = \sum_{\lambda=1}^{1} b_{\lambda S}^2 / \lambda_{\lambda S}^{(2)}.
\]

Given estimates of the \( a_{\lambda S} \) and \( b_{\lambda S} \), estimates of the wind, vorticity and divergence are obtained analytically for any desired \( P \). The resulting estimated wind field is called a vector spline on the sphere. \( \lambda \) is the bandwidth parameter, it controls the partitioning of the data vector \((u_1, \ldots, u_n, v_1, \ldots, v_n)\)' into a part due to "signal" and a part due to noise. \( \delta \) can be viewed as a signal partitioning parameter, it controls the partitioning of the "signal" part of the data into a divergent part and a non-divergent part.

A Monte Carlo study was carried out to test the effectiveness of the method and to determine whether good estimates of \( \lambda \) and \( \delta \) could be obtained from the data by GCV. The results are reported in Wahba (1982), and in preparation. The weights \( \lambda_{\lambda S}^{(1)} \) and \( \lambda_{\lambda S}^{(2)} \) were adapted from the data collected by Stanford (1979). Realistically scaled model 500 mb wind fields were generated via a Karhunen-Loeve expansion in stream function and velocity potential, the resulting "true" wind vectors at 114 North American weather stations computed and a random 2.5 m/sec rms measurement error added to each computed wind component. Good recovery of winds, vorticity and divergence in an area covered by the data grid and extending a small amount past it was obtained using the estimated \( \lambda \) and \( \delta \). (The individual \( a_{\lambda S}, b_{\lambda S} \) are not recovered from only North American data). The results are sensitive to both \( \lambda \) and \( \delta \). It can be seen from the experiments that a poor value of \( \delta \) causes obvious edge effects and that fixing \( \delta \) at values which oversuppress the divergent part of the wind tend to cause increased errors in the estimation of the non-divergent part.

4. ON A SINGLE VARIATIONAL PROBLEM FOR MERGING DATA, FORECAST AND PRIOR KNOWLEDGE OF ATMOSPHERIC SPECTRAL ENERGY DISTRIBUTION

We claim that a set of "moderately" reasonable assumptions concerning the forecast errors and measurement errors, along with prior knowledge concerning the spectral energy distribution in the atmosphere, leads to an "optimal"
initial state obtained as the solution of a single (large) variational problem. This variational problem will have the same number of "unknowns" as the degrees of freedom in the forecast model, thus raising questions of numerical feasibility. We brush aside computational questions for the present, adopting the point of view that it is worthwhile to examine an "optimal" variational problem, and then ask, how close can one come to computing a reasonable approximation to it. (Some numerical shortcuts appear in Bates and Wahba (1982)). One of the advantages of examining the variational form is that it is fairly evident how to add side constraints based on the physics.

A (simplified) example goes as follows. Consider a spectral model where the state of the atmosphere is expressed in terms of Hough functions. Using notation similar to Tribbia (1982), and considering a single level, let

\[
\begin{align*}
(U, V, \phi) &= \sum_{j=1}^{N_{\text{max}}} X_j H_j^R + \sum_{k=1}^{M_{\text{max}}} Y_k H_k^G \\
&= (X_1^F, \ldots, X_{N_{\text{max}}}^F; Y_1^F, \ldots, Y_{M_{\text{max}}}^F) 
\end{align*}
\] (4.1)

(U, V, \phi) is the "true" wind and geopotential field, and H_j^R and H_k^G are rotational (slow) and gravitational (fast) Hough functions. We suppose that the true "state" of the atmosphere is adequately described by the N_{\text{max}} + M_{\text{max}} vector \theta = (\chi; \psi) = (X_1, \ldots, X_{N_{\text{max}}}; Y_1, \ldots, Y_{M_{\text{max}}}) and an analysis consists of obtaining an updated estimate \theta of \theta, given a forecast

\[
\theta^F = (X_1^F; Y_1^F) = (X_1^F, \ldots, X_{N_{\text{max}}}^F; Y_1^F, \ldots, Y_{M_{\text{max}}}^F),
\] (4.2)

given data (u_i, v_i, \phi_i) representing observations on the wind and geopotential height at point P_i, and given the "prior knowledge" concerning atmospheric energy distribution obtained, e.g. from data like that collected by Kasahara and Puri (1981). This prior knowledge is of the form

\[
\begin{align*}
E_X j &= b_1 \lambda_j^R \\
E_Y k &= b_2 \lambda_k^G
\end{align*}
\] (4.3)

where we assume that the relative energies \lambda_j^R and \lambda_k^G are known but that b_1 and b_2 are not. We will assume that all cross covariances E X_i X_j, E X_i Y_j and E Y_i Y_j can be approximated by 0. Suppose further that

\[
\begin{align*}
E(X_j - X_j^F)(X_k - X_k^F) &= \omega_{\chi \chi} \sigma_{jk} \\
E(Y_j - Y_j^F)(Y_k - Y_k^F) &= \omega_{\psi \psi} \sigma_{jk} \\
E(X_j - X_j^F)(Y_k - Y_k^F) &= \omega_{\chi \psi} \sigma_{jk}
\end{align*}
\] (4.4)
where the $\sigma_{jk}$'s are known, and let $\Sigma$ be the $(N_{\text{max}} + M_{\text{max}}) \times (N_{\text{max}} + M_{\text{max}})$ covariance matrix with entries $\sigma_{jk}^{XX}$, $\sigma_{jk}^{XY}$, $\sigma_{jk}^{YY}$. Let the measurement errors of $(u_i, v_i, \phi_i)$ be independent with variances $\omega_0 \sigma_u$, $\omega_0 \sigma_v$, $\omega_0 \sigma_\phi$. Suppose all random variables are normally distributed. Then

**Theorem**: (G. Wahba, in preparation) The Bayes estimate $\hat{\theta} = (\hat{X} : \hat{Y})$ of $\theta = (X : Y)$ is the minimizer of

$$
\frac{1}{n} \sum_{i=1}^{n} \left\| \left( \frac{u_i}{v_i} \right) - \frac{N_{\text{max}}}{\Sigma} \sum_{j=1}^{N_{\text{max}}} X_j R_j(P_i) - \frac{M_{\text{max}}}{\Sigma} \sum_{k=1}^{M_{\text{max}}} Y_k H_k(P_i) \right\|^2 \sigma
$$

$$
+ \omega \left( (X - X^F : Y - Y^F) \right) \sum^{-1} (X - X^F : Y - Y^F)'
$$

$$
+ \lambda \left( \sum_{j=1}^{N_{\text{max}}} \frac{X_j^2}{\lambda_j^R} + \delta \sum_{k=1}^{M_{\text{max}}} \frac{Y_k^2}{\lambda_k^G} \right)
$$

where $\left\| \begin{pmatrix} u \\ v \\ \phi \end{pmatrix} \right\|^2 = \frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} + \frac{\phi^2}{\sigma_\phi^2}$, and $\omega = \omega_0 / \omega_\text{fn}$, $\lambda = \omega_0 / b_1 n$, $\delta = b_1 / b_2$.

Since the expression to be minimized is a quadratic form in the components of $(X : Y)$, (for given $\omega$, $\lambda$, $\delta$), the minimizer can be readily expressed as the solution to a linear system. It is conjectured that the bandwidth parameter $\lambda$, the partitioning parameter $\delta$ and the "error balancing" parameter $\omega$, can all be estimated (simultaneously!) from the data by GCV. Note that a large $\omega$ suppresses forecast relative to observational data and a large $\delta$ suppresses fast modes relative to slow modes. The results reported in Section 3 suggest that at least one bandwidth and one (appropriately chosen) partitioning parameter can be chosen by GCV. Part of the art of this approach is to choose these "tuning" parameters so that they are (a) the important ones and (b) their determination is relatively "well posed." Sensitivity of optimum interpolation to bandwidth parameters has recently been discussed by Hollingsworth (1982), Lorenc (1981) and others.

To be practical it may be necessary to assume an oversimplified structure for $\Sigma$ and/or to break up the variational problem into a series of sub-problems. The usual Kalman filtering theory would give $\theta$ as the minimizer of the above expression, with $\lambda = 0$ and with $(\omega E)$, the forecast error covariance, given by a recursion formula in time. Our description above,
implicitly assumes the existence of a limiting \( w_F^E \). The forecast of Ghil et al. would, roughly speaking reduce in this context, to setting \( \lambda = 0 \) and constraining \( Y \) to be 0. The above theorem generalizes to the simultaneous analysis of all levels. Then satellite radiance data (which "cuts across" all levels) can be included as another term in the variational formulation. Physical constraints can, in principle be included as side conditions. It also appears that non-linear balancing constraints related to those discussed in, e.g. Tribbia (1982), whose purpose is to minimize the propagation of unwanted gravity waves in the non-linear forecast equations, may also be incorporated as either weak or strong constraints.

ACKNOWLEDGMENTS

This research was supported by NASA Grant NAG5-128.

REFERENCES


Hollingsworth, A., (1982) Operational data assimilation at ECMWF, these proceedings.


Phillips, N. A., (1982a) A very simple application of Kalman filtering to meteorological data assimilation, Office note 258, NMC.


Because of a mis-match between analyses and forecast models, numerical forecasts often exhibit high frequency noise. Nonlinear normal mode initialization successfully eliminates these spurious high-frequency oscillations from the forecasts (Daley, 1979; Williamson and Temperton, 1981) but, in doing so it often produces modifications to the analyses which exceed the expected analyses error even over data-rich regions such as the continental United States. These changes are in part due to short comings of both aspects (analysis and initialization) of the system. The analysis has errors due to, for example, adopting a linear geostrophic assumption while the model (and atmosphere) actually satisfy a more complicated nonlinear gradient relationship (Williamson, et al., 1982). The unconstrained nonlinear normal mode initialization does not recognize the specific nature of this analysis error and thus does not treat it properly.

The variational approach of Daley (1978) and Tribbia (1982) was developed to address this problem. It minimizes a weighted difference between the balanced initial state produced by nonlinear normal mode

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.
initialization and an analysis. The weights are related to the confidence in the analysis at each point. However, a more desirable approach might be to minimize the difference between the balanced initial state and the observations themselves rather than an analysis of these observations.

Optimal interpolation can be looked at as a minimizational fit to the observations. In fact, Kimeldorf and Wahba (1970) show that for every covariance structure function in optimal interpolation, there is a variational problem that will give the same solution and vice versa. Therefore, in Tribbia's (1982) iterative variational approach, it seems reasonable to replace his variational minimization step with optimal interpolation. In other words, one could perform optimal interpolation followed by nonlinear normal mode initialization followed by a second optimal interpolation (using the first initialized analysis for the first guess or trial field), followed by initialization, etc., until, hopefully, the initialization and the analysis both make negligible changes, indicating convergence. Of course, sequential optimal interpolation would offer no improvement over the first analysis if nothing was done in between, since no new information is added and the first optimal interpolation would minimize the expected analysis error within that formalism. However, the slow component of the analysis generally has a smaller error than the first trial field and therefore the subsequent geostrophic analysis increments will be smaller in the next optimal interpolation, resulting in smaller error from the geostrophic approximation. In such an iterative approach on successive iterations one could also introduce local variations in the structure functions which are
derived from the previous iterate. We will not pursue that possibility further here, but rather concentrate on the geostrophic error.

Following Tribbia's (1982) examples (especially his Fig. 5) the iterative procedure introduced above can be interpreted geometrically using the slow manifold diagram introduced by Leith (1980) and further elaborated by Daley (1980). Figure 1 illustrates the procedure for an idealized case. The abscissa R is the Rossby manifold, the ordinate G the gravity manifold, the curve S is the slow manifold, and the lines D are data manifolds for a particular type of data. For the purpose of illustration, we take D to consist of height observations with no errors and further assume that the optimal interpolation scheme draws to the data exactly so the height analysis has no error.

Figure 1
Optimal interpolation requires a first guess or trial field which is modified according to the observations. In our iterative approach there is a sequence of trial fields. The first trial field for the iterative procedure would normally be provided by a model forecast as is generally done in operational practice today and thus lies on (or can be made to lie on) the slow manifold and is denoted $S_0$ in Fig. 1. The optimal interpolation with geostrophically related covariances modifies the Rossby modes without significantly altering the gravity modes of the trial field to produce the first analysis $A_1$ in Fig. 1, which under the above assumption lies on the height data manifold $D$. The analysis procedure itself is then represented by a horizontal line denoted $OI$ in the Figure. Nonlinear normal mode initialization modifies the gravity waves of the analysis $A_1$ to produce its slow component $S_1$ lying on the slow manifold. The initialization procedure is represented by a vertical line in Fig. 1 and denoted by NMI. A second analysis is then performed based on $S_1$ for the (second) trial field to produce the second analysis $A_2$ after which its slow component $S_2$ is calculated. The procedure is continued with the slow component of the $n^{th}$ analysis serving as the trial field for the $(n+1)^{st}$ analysis. Convergence is indicated by insignificant changes being made by both the analysis and the initialization. Figure 1 indicates that for cyclonic flows the error of the slow component height field changes sign with each analysis since it lies on the opposite side of the data manifold that the previous one, but for anticyclonic flows the error of the slow component height field always has the same sign as the error of the first guess while its magnitude decreases. Both cases are seen to converge in this simplified description.
The above example illustrates the procedure for an idealized case which will never exist in reality. Observations are not error-free and they are not collocated with analysis points so the analysis will have additional errors other than the geostrophic one which might affect the convergence properties of the iterative approach. We have tested this approach in a more practical situation with a few simplifications to isolate the convergence of properties of the iterative approach from other problems associated with nonlinear normal mode initialization itself. Results of these tests are presented in Williamson and Daley (1982). They show that the iterative optimal interpolation - nonlinear normal mode initialization provides an analysis consistent with the observational errors and in nonlinear balance. The errors associated with geostrophically related covariances are eliminated. This occurs because the slow component of each successive analysis is a better approximation to the atmospheric state than the trial field used for that analysis. Thus, when the slow component is used as the trial field for the next analysis, the analyzed geostrophic increments are smaller and the geostrophic component of the analysis is smaller until eventually it is negligible.

The iterative approach need not be prohibitively expensive as successive analyses need only be performed in regions which are not in balance, rather than over the whole domain. The fast component of the analysis provides a local and global measure of convergence. In regions where it is small, the procedure has converged and the analysis need not be repeated. It is only in regions with a relatively large fast component
that further analyses will be beneficial. Given height data only, these regions are limited to areas of large curvature which generally cover only a small fraction of the globe.

REFERENCES


What I Learned at the 1982 Stanstead Seminar

Well, you see, for two make your "Numbing Weather Predilections" (NWP, for short) your needs two do things too the obfuscations that are coming from around the world.

The first thing is your "Opti-Mum Interogation" (OI, for short). This is where you throw your obfuscations into your statical meat grinder two get rid of your gross errors. They calls it "Opti-Mum" because the people who use it are optimists and they keeps quiet about its problems.

The second thing is your "Normal Mood Initials" (NMI, for short). This is where you fiddle your obfuscations to put back your gross errors. They're too NMI's that people couldn't make up they're minds. You have your "Machenhowitzer" that shot down Walker's circulation and your "Bare-Trivial" that was trivial two do and did barely nothing.

The best I learned was really your OI and your NMI. This means the weatherpersons say 0 I Need More Information.

Signed

C. Farquarson

Charlie Farquarson