BAB

MET O 11 TECHNICAL NOTE NO 168

THE OPERATIONAL DATA ASSIMILATION SCHEME

by

W H LYNE, C T LITTLE, R K DUMELOW and R S BELL (Met O 2b)

1st Edition

March 1983

Met O 11 (Forecasting Research Branch)
Meteorological Office
London Road
Bracknell
Berkshire
U.K.

CONTENTS

DEFINITION OF VARIABLES AND SYMBOLS IN SECTIONS 1 TO 4

SECTIONS

1. INTRODUCTION

2. GENERAL PRINCIPLES

3. ANALYSIS

- 3.1 General description
- 3.2 Observational errors
- 3.3 Background field errors
- 3.4 Mode 1 quality control
- 3.5 Selection of data for optimal interpolation
- 3.6 Observational error correlations
- 3.7 Background field error correlations
- 3.8 Mode 2 quality control
- 3.9 Calculation of the weights

4. ASSIMILATION

- 4.1 General description
- 4.2 Interpolation to analysis grid points
- 4.3 Interpolation to model grid points
- 4.4 Specification of relaxation coefficients
- 4.5 Geostrophic wind increments
- 4.6 Hydrostatic temperature increments
- 4.7 Divergence damping

CONTENTS

DEFINITION OF VARIABLES AND SYMBOLS IN SECTIONS 1 TO 4

SECTIONS

1. INTRODUCTION

2. GENERAL PRINCIPLES

3. ANALYSIS

3.1 General description

3.2 Observational errors

3.3 Background field errors

3.4 Mode 1 quality control

3.5 Selection of data for optimal interpolation

3.6 Observational error correlations

3.7 Background field error correlations

3.8 Mode 2 quality control

3.9 Calculation of the weights

4. ASSIMILATION

4.1 General description

4.2 Interpolation to analysis grid points

4.3 Interpolation to model grid points

4.4 Specification of relaxation coefficients

4.5 Geostrophic wind increments

4.6 Hydrostatic temperature increments

4.7 Divergence damping

REFERENCES

FIGURES

ANNEXES

1. Design of the analysis program

2. Design of the assimilation program

3. Data fields

4. Grids

5. Finite difference approximations

6. Observational error correlations

7. Background field error correlations

8. Observational error levels

9. Background field error levels

10. Current values of analysis and assimilation parameters

DEFINITION OF VARIABLES AND SYMBOLS IN SECTIONS 1 TO 4

1. General

x_{α}^{β}	general atmospheric variable
p	pressure
T	temperature
Θ	potential temperature
u&v	westerly and southerly wind components
r	specific humidity
R	gas constant
C_p	specific heat of atmosphere at constant pressure
K	ratio R/C_p (but alternative specification of relaxation coefficient in Section 1)
f_{α}	Coriolis parameter

2. Analysis

sub-script α	i&j	observation points
	g	grid point
	s	surface variable
super-script β	a	analysed or optimally interpolated value
	b	background value
	o	observed value
	t	true value
	$\emptyset$	geopotential height
	W_{α}	optimal interpolation weight
	E_{α}^{β}	root mean square expected error
	$E(x_{\alpha})$	root mean square expected error for variable x_{α}
	$\varepsilon_{\alpha}^{\beta}$	error i.e. $x_{\alpha}^{\beta} - x_{\alpha}^{t}$
	$\langle\ \rangle$	ensemble mean
	n_1	number of standard deviations of error allowed in Mode 1 quality control

n_2 number of standard deviations of error allowed i[n] [M]ode 2 quality control

Ri radius of influence defining horizontal search area

R_p pressure ratio defining vertical search extent

M_1 maximum number of observations selected for each grid point position

M_2 maximum number of data selected for each grid point variab[le]

M_3 maximum number of pressure levels of each complete observation from which data may be selected for each grid point variable

F adjustment factor for variables in data selection

$\mu_{\alpha\gamma}^{\beta}$ error correlation function

$\mu(x_\alpha, y_\gamma)$ model error correlation function for variables x and y

a&b parameters defining model height error correlation function

c&d parameters defining model humidity error correlation function

p&q parameters defining satellite temperature error correlation function

$r_{\alpha\gamma}$ horizontal distance between points α and γ

$\alpha_{\alpha\gamma}$ angle between line of latitude at point α and the line joining point α to point γ

3. Assimilation

sub-script α (t time
(k model level

$\underline{x}_\alpha$ vector of model variables

λ relaxation coefficient

Δx_α assimilation increment for variable x

Δt time step

Δs grid length

Φ geopotential height

α relaxation coefficient for hydrostatic temperature increments

β relaxation coefficient for geostrophic wind increments

$H(\Delta x_\alpha)$ hydrostatic temperature increment

$G(\Delta x_\alpha)$	geostrophic wind increment
$A(x_\alpha)$	adjustment term of model equations
$B(x_\alpha)$	advection term of model equations
$\frac{\partial}{\partial\lambda}$	partial **derivative** with respect to longitude
$\frac{\partial}{\partial\phi}$	partial derivative with respect to latitude
$\cos\phi$	cosine of latitude
a	radius of Earth
π	$(Ps/1000)^K$
μ	divergence damping coefficient
D_d	divergence damping term in model equations
σ	P/Ps vertical co-ordinate of model
D	divergence
ζ	vorticity
C	phase speed of fastest moving gravity wave

1. INTRODUCTION

The new operational data assimilation scheme has evolved from a technique first suggested by Lorence (1976) in the context of Observing System Simulation Experiments. Observational data, or rather their differences from model values, are interpolated to model grid points, and the resulting increments are used repeatedly to correct the model as it is integrated over a so-called assimilation period. This technique is in fact analogous to the dynamical relaxation methods of Hoke and Anthes (1976) and Davies and Turner (1977), and is very similar to a form of data assimilation described by Miyakoda et al (1978).

An opportunity to assess the method in a semi-operational environment was furnished in 1979 by the First GARP Global Experiment (FGGE). An experimental data assimilation scheme was constructed and, during the two Special Observing Periods, global analyses of all model variables were produced at six-hourly intervals. The scheme is described by Birch and Lyne (1980) and some of the findings from the experiment are given by Lyne et al (1982). The results were on the whole encouraging, and the decision was made to base the next operational data assimilation scheme on the system written for FGGE, drawing on the experience gained during the experiment.

2. GENERAL PRINCIPLES

The data assimilation scheme consists of two basic stages. The first, or analysis stage, consists of the calculation of weights to interpolate the observations to model grid points, and incorporates the automatic quality control of the data. The interpolation of data has been found to be a very necessary requirement for its adequate assimilation (Ghil et al 1979), and Lyne (1978) has shown it to be more important than the second or assimilation stage. This consists of using the interpolation weights to calculate corrections to the model values at each time step of a model forecast. These corrections are scaled by a factor λ and then added into the model fields. Thus a value of $\lambda = 0$ corresponds to a model forecast, whilst $\lambda = 1$ for one time step corresponds to a conventional analysis. The correct value of λ can only be adequately determined by experiment and experience since the optimal choice depends on the data available and the scale of the meteorological feature being analysed. Too low a value will lead to a poor fitting of observations, whilst too high a value may excite large amplitude gravity-inertial "noise". Lyne (1979) has analysed the process in a simplified system and has shown that the factor λ can be related to the relaxation coefficient K of Davies and Turner (1977) and Hoke and Anthes (1976) by

$$\lambda = K\Delta t/(1+K\Delta t)$$

where Δt is the time step of the model.

In the new operational scheme, data are assimilated every six hours with all data falling within three hours of the analysis time being assigned to that time. These data are used to corect the model as it is integrated over the six hour period from the previous analysis time, the resulting fields being the initial fields for a model forecast or the next assimilation cycle. This system is depicted in Figure 2.1.

The analysis and assimilation stages are described in more detail in the following sections. There are many disposable parameters in the scheme whose values may change from time to time. In order that this note may be kept up to date more easily their current values are specified in the Annexes rather than in the text.

Finally mention should be made of two problems which, although not the concern of this note, give rise to operational difficulties. Firstly over areas of high or rapidly changing topographic height, the model surface height may differ quite considerably from that of an observing station. Surface data have then to be extrapolated to the model's surface, and low-level data may be used at inapporopriate levels. This can give rise to significant errors. Secondly the analysis and assimilation have to handle files which, as well as being large, also have substantial variations in their record lengths. Thus the file of operational interpolation weights can be of the order of 3000 sectors long (1 sector is 512 full 64 bit words), whilst individual records can vary between 1 sector and lengths of the order of 70 sectors. This can lead to inefficiencies in the use of computer storage and I/O bound stages in processing.

3 ANALYSIS

3.1 General Description

In the first or analysis stage, a method known as optimal interpolation is used to calculate weights for the interpolation of observations to model grid points (Gandin 1963, Rutherford 1972). Differences between observed values and a background or first guess field are calculated at observation points and are then linearly interpolated to model grid points.

$$x_g^a - x_g^b = \sum_i W_i \, (x_i^o - x_i^b) \tag{3.1}$$

Here x is the variable to be interpolated, subscripts g and i refer to grid and observation points respectively, and superscripts a, b and o refer to analysed, background and observed values. Although equation (3.1) is univariate, optimal interpolation in its most general form can be multivariate with the analysed variable being of a different type to the observations (Lorenc 1981).

The analysis is optimal in the sense that the weights W_i are chosen so that the expected error variance ${E_g^a}^2$ of the interpolated value is a minimum where

$$\begin{aligned} {E_g^a}^2 &= \langle (x_g^a - x_g^t)^2 \rangle \\ &= \langle {\varepsilon_g^a}^2 \rangle \\ &= \langle [\varepsilon_g^b + \sum_i W_i (\varepsilon_i^o - \varepsilon_i^b)]^2 \rangle \end{aligned} \tag{3.2}$$

A superscript t denotes the true value, ε the error and $\langle \; \rangle$ an ensemble mean. It can be shown that the W_i must satisfy the following set of simultaneous equations, the number of equations being equal to the number of observations used in interpolating to a model grid point.

$$\sum_j W_j \{ \langle \varepsilon_j^o \, \varepsilon_i^o \rangle + \langle \varepsilon_j^b \, \varepsilon_i^b \rangle - \langle \varepsilon_j^o \, \varepsilon_i^b \rangle - \langle \varepsilon_i^o \, \varepsilon_j^b \rangle \}$$

$$= \langle \varepsilon_i^b \, \varepsilon_g^b \rangle - \langle \varepsilon_i^o \, \varepsilon_g^b \rangle \qquad (3.3)$$

On the assumption that observations are unbiased and not correlated with the background values, equations(3.3) reduces to

$$\sum_j W_j \{ \langle \varepsilon_j^o \, \varepsilon_i^o \rangle + \langle \varepsilon_j^b \, \varepsilon_i^b \rangle \} = \langle \varepsilon_i^b \, \varepsilon_g^b \rangle \qquad (3.4)$$

where $\langle \varepsilon_j^o \, \varepsilon_i^o \rangle$ and $\langle \varepsilon_j^b \, \varepsilon_i^b \rangle$ are the observational and background field error covariances respectively. The expected analysis error may then be shown to be

$$E_g^{a^2} = E_g^{b^2} - \sum_i W_i \langle \varepsilon_i^b \, \varepsilon_g^b \rangle \qquad (3.5)$$

where $E_g^{b^2} = \langle \varepsilon_g^{b^2} \rangle$ is the expected mean square error of the background field.

The quality control of data is performed in two stages. In the first, or Mode 1 check, data are compared to the model background field and any observations differing by more than a critical amount from the expected error are flagged, that is if

$$(x_i^o - x_i^b)^2 > n_1^2 [E_i^{o^2} + E_i^{b^2}] \qquad (3.6)$$

For wind data the vector wind error is the quantity checked.

In the second, or Mode 2 check, the observation is compared with a value interpolated from surrounding data, and a similar criterion is used to determine whether it is flagged.

$$(x_i^o - x_i^a)^2 > n_2^2 [E_i^{a^2} + E_i^{o^2}] \qquad (3.7)$$

In calculating the interpolated value x_i^a, neither the observation itself, nor data flagged in Mode 1 or by the synoptic data bank, is used. This Mode 2 check is the principal arbiter as to whether data is accepted or rejected, though there is a facility for it to be overruled by the specificiation of a flag in the initial data sets. Note that although an observation may be flagged in Mode 1, it can still be accepted if it passes the Mode 2 check.

Aspects of the analysis stage are described in more detail in the following subsections, and an account of the principles used in the design of the computer program and its structure is given in Annex 1. The formats of the data files processed during the analysis are given in Annex 3.

3.2 Observational Errors

The root mean square observational errors E^o are specified in a look-up table and vary only with respect to observation type and with height. Values are specified on standard pressure levels and there is no interpolation in the vertical; the value taken is that of the standard pressure level nearest the observation. Values of the error levels currently used are given in Annex 8.

3.3 Background field errors

The root mean square background field errors E^b are carried in a separate data file and are specified on a coarse 5^o latitude-longitude grid and on six pressure levels in the vertical. Values are calculated at observation points and analysis grid points by bi-linear interpolation in the horizontal and by linear interpolation in log (pressure) in the vertical. Values on pressure levels outside the range of those of the error data set are found by extrapolation. Some of the values being currently used are given in Annex 9.

3.4 Mode 1 quality control

In order to perform both quality control checks, values of the background fields need to be calculated at observation points. The background field is, in practice, the latest available model forecast valid at the analysis time, and this is a six hour forecast from the previous assimilation. Because the model grid has a high resolution (see section 4), a sufficiently accurate value can be calculated by using the same interpolation routines employed for the background errors, ie bi-linear interpolation in the horizontal and linear-in log (pressure) in the vertical. The differences between the observations and these interpolated values can then be calculated, and, with the previously calculated observational and background field errors, the check in expression (3.6) can be made.

Values of n_1 depend on the observation type, those being more liable to gross errors, that is errors which do not belong to the normal error distribution, being assigned smaller values than those less likely to suffer from this defect. An example of the former is an aircraft report, whilst a radio-sonde ascent is an example of the latter. Current values of n_1 are given in Annex 10.

3.5 Selection of data for optimal interpolation

Ideally all data should be used to interpolate to each grid point, this being most important for multivariate analysis when the interpolated increments $(x^a - x^b)$ can be made to obey constraints such as geostrophy and non-divergence in the wind field. Lorenc (1981) has described a scheme which approximates to this ideal, but, in general, computing considerations dictate that some form of selection of the best data must be made for each grid point. This is particularly true in the context of this data assimilation scheme, otherwise the file of interpolation weights generated for the assimilation stage would be far too large. (See remarks in section 2).

Data are selected in two stages: firstly the nearest observations falling within a certain radius of influence (Ri) are chosen, up to a given maximum number (M_1); secondly, for each variable at each level of each grid point, the best data are chosen up to a maximum number(M_2). In this context an observation is a reported observation such as an upper air ascent which may contain many data e.g. wind, temperature and humidity data at several levels in the vertical.

In the first stage of selection the horizontal distances between the observations and grid points are calculated, and are adjusted by a factor F of the radius of influence before the search for the nearest is made.

$$\text{adjusted distance} = \text{actual distance} + F \times Ri$$

The factor F is a function of observation type, and the adjustment is made to ensure that the more valuable multi-level observations, such as upper air ascents, are selected despite the presence of many nearer single level observations such as aircraft reports.

In the final stage of selection, the best data are defined as those which, when taken singly, reduce the expected interpolation error E_g^a by the greatest amount. However, to ensure an even distribution around a grid point, at least one item of data is chosen from each quadrant around the grid point, provided such an item is present. In addition, to maximise the number of independent pieces of information, a limit is placed on the number of pressure levels (M_3) from which data may be selected from each complete observation. These pressure levels must lie within a certain distance of the grid point level, this distance being defined in terms of a ratio R_p. Thus for selection

$$|\ln(p_i/p_g)| < \ln R_p$$

The current values of all these parameters may be found in Annex 10.

3.6 Observational error correlations

The observational error covariances in equation (3.4) are expressed as products of root mean square errors and correlations

$$\langle \varepsilon_j^o \; \varepsilon_i^o \rangle = E_j^o \; \mu_{ji}^o \; E_i^o \tag{3.8}$$

For most types of observations the error correlations are zero, only being significant for several observations of a variable from a single instrument e.g.

temperature errors from an upper air ascent and errors in satellite deri)
temperatures from a single instrument.

Observational error correlations are specified as analytic functions of distance, and the observations which are currently assumed to have significant error correlations are given in Annex 6 together with the corresponding functions.

3.7 Background field error correlations

The background field error covariances in equation (3.4) are also expressed as products of root mean square errors and correlations

$$\langle \varepsilon_j^b \, \varepsilon_i^b \rangle = E_j^b \, \mu_{ji}^b \, E_i^b \tag{3.9}$$

Analytic functions are again specified for the correlations but are based on a single height-height error correlation function

$$\mu(\phi_i, \phi_j) = \exp\left(-a r_{ij}^2 - b \ \ln^2(p_i/p_j)\right) \tag{3.10}$$

where r_{ij} is the horizontal distance from the i^{th} to the j^{th} observation. This function was chosen because it exhibits the correct qualitative behaviour of the expected error distribution, and because it leads to a positive definite system of simultaneous equations (3.4) ensuring non-singularity. A single base function was used so that constraints such as geostrophy could be enforced in deriving the error covariances for full multivariate analysis (Lorenc 1981). The derivation of these functions is given in Annex 7.

The auto-correlations of the primary model variables p_s (surface pressure), θ (potential temperature), u and v (westerly and southerly wind components) and r (specific humidity) are given below.

$$
\left.
\begin{aligned}
&\mu(p_{si},p_{sj}) = \mu(\phi_i,\phi_j) \\
&\mu(\theta_i,\theta_j) = \left[1 - 2b\ \ln^2(p_i/p_j)\right] \mu(\phi_i,\phi_j) \\
&\mu(u_i,u_j) = \cos\alpha_{ij}\ \cos\alpha_{ji}\left[1 + \tan\alpha_{ij}\ \tan\alpha_{ji}\quad (1 - 2ar_{ij}^2)\right] \mu(\phi_i,\phi_j) \\
&\mu(u_i,v_j) = \cos\alpha_{ij}\ \cos\alpha_{ji}\left[\tan\alpha_{ji} - \tan\alpha_{ij}\quad (1 - 2ar_{ij}^2)\right] \mu(\phi_i,\phi_j) \\
&\mu(v_i,u_j) = \cos\alpha_{ij}\ \cos\alpha_{ji}\left[\tan\alpha_{ij} - \tan\alpha_{ji}\quad (1 - 2ar_{ij}^2)\right] \mu(\phi_i,\phi_j) \\
&\mu(v_i,v_j) = \cos\alpha_{ij}\ \cos\alpha_{ji}\left[\tan\alpha_{ij}\ \tan\alpha_{ji} + \quad (1 - 2ar_{ij}^2)\right] \mu(\phi_i,\phi_j) \\
&\mu(r_i,r_j) = \exp\left(-cr_{ij}^2 - d\ \ln^2(p_i/p_j)\right)
\end{aligned}
\right) \quad (3.11)
$$

where α_{ij} is the angle between the line of latitude at point i and the line joining point i to point j. The covariances are formed by multiplying the correlations by the relevant root mean square error values, but to enforce hydrostatic balance in the interpolated increments these need to be related to the height errors as follows:

$$E(p_{si}) = E(\phi_i)\frac{p_{si}}{RT_{si}} \quad (3.12)$$

$$E(\theta_i) = E(\phi_i)\left(\frac{1000}{p_i}\right)^K \frac{\sqrt{2b}}{R} \quad (3.13)$$

where T_s is the surface temperature, R is the gas constant and K is the ratio of R to the specific heat at constant pressure C_p.

For geostrophy the wind errors have to satisfy

$$E(u_i) = E(v_i) = E(\phi_i)\ \frac{\sqrt{2a}}{f_i} \quad (3.14)$$

where f_i is the value of the Coriolis parameter at point i

Although the wind and temperature errors in the background field error file are specified independently, their values in middle and high latitudes can be adjusted through these relations to satisfy the constraints either exactly or approximately.

The parameters a, b, c and d which define these functions are themselves functions of position, and current values are given in Annex 10. However, in calculating the weights for any one grid point variable their values remain constant for each of the selected observation points. If this were not so, then the necessary condition of positive definiteness in the simultaneous equations could not be guaranteed.

3.8 Mode 2 quality control

With the observation to be checked taking the place of the grid point, equations (3.4) are solved for the interpolation weights W_i and expression (3.7) constructed. The values of n_2 again depend on observation type for the same reasons given for the Mode 1 values, and the current specification is given in Annex 10. All data are subject to this check, but only data which passed the Mode 1 checkand were not flagged by the data bank are used in calculating the interpolated value x_i^a.

3.9 Calculation of the weights

The equations (3.4), which are symmetric positive definite, are solved by Gaussian elimination, and the resulting weights written to a file for their subsequent use during the assimilation stage (see Annex 3 for the format of this file). The weights are calculated at the grid points of the analysis grid which do not necessarily coincide with those of the background or model grid. The current analysis grid is given in Annex 4.

ASSIMILATION

4.1 General description

In the final data assimilation stage the data, or rather their differences from model values, are interpolated to analysis grid points using the optimal interpolation weights calculated during the analysis stage. These differences are then interpolated to the model grid and are used to correct the model at each time step of a six hour forecast ending at the analysis time (see Figure 2.1). Annex 4 contains details of both analysis and model grids. The integration scheme of the model is described in more detail in a companion note (Temperton et al 1983 henceforth referred to as FC), but for our purposes it is sufficient to note that it is split explicit with three adjustment steps to each full advection step. The corrections are calculated once for each advection time step, scaled by a factor λ and the resulting increments added in at each of the adjustment time steps. Further increments based on geostrophic and hydrostatic assumptions are calculated from these corrections in order to hasten the adjustment process during the assimilation.

The assimilation equations can be represented schematically by

$$\left.\begin{aligned}
\underline{\Delta x} &= \sum_i W_i (x_i^o - x_t) \\
\underline{x}'_{t+\Delta t/3} &= \underline{A}(\underline{x}_t) + \lambda \underline{\Delta x} + \alpha \underline{H}(\underline{\Delta x}) + \underline{D}_d(\underline{x}_t) \\
\underline{x}'_{t+2\Delta t/3} &= \underline{A}(\underline{x}'_{t+\Delta t/3}) + \lambda \underline{\Delta x} + \alpha \underline{H}(\underline{\Delta x}) + \underline{D}_d(\underline{x}'_{t+\Delta t/3}) \\
\underline{x}'_{t+\Delta t} &= \underline{A}(\underline{x}'_{t+2\Delta t/3}) + \lambda \underline{\Delta x} + \alpha \underline{H}(\underline{\Delta x}) + \underline{D}_d(\underline{x}'_{t+2\Delta t/3}) \\
\underline{x}_{t+\Delta t} &= \underline{B}(\underline{x}'_{t+\Delta t}) + \beta \underline{G}(\underline{\Delta x})
\end{aligned}\right\} \qquad (4.1)$$

where Δx are the corrections calculated from the observed values x_i^o and the model values at time t using the optimal interpolation weights W_i, and an under-bar denotes a vector over model points. A prime denotes an intermediate value during the full advection time step Δt and the operators A and B represent the adjustment and advection terms of the forecast equations respectively. The term B also contains the physical parametrizations. The additional operators H and G represent the hydrostatically derived temperature increments and the geostrophically derived wind increments respectively. Gravity wave noise generated by local imbalances during the assimilation process could cause initial roughness in the subsequent forecast and this in turn could lead to the rejection of good data in a subsequent quality control step. This is controlled both by the standard non-linear diffusion of the model (see FC) and by the addition of a form of divergence damping represented by the operator D_d. Current values of diffusion coefficients and of the scaling factors λ, α and β are given in Annex 10. These processes are described in more detail in the following subsections together with other aspects of the assimilation stage.

The final field of assimilation increments are Fourier truncated near the Poles to remove unstable modes. The companion note on the forecast model (FC) describes this procedure and the reasons why it is necessary.

An account of the principles used in the design of the computer program and its resulting structure is given in Annex 2. Annex 3 contains details of the data files presented to the assimilation apart from the model fields themselves. These are detailed in FC.

4.2 Interpolation to analysis grid points

At each full advection time step the optimum interpolation equation (3.1) is used to calculate corrections at each analysis grid point. In this assimilation stage the background value is replaced by the latest available model value but otherwise the calculation is identical. Model values are interpolated bi-linearly in the horizontal to observation points, and linearly in log (pressure) in the vertical to the data levels. The difference between observed and model values can then be calculated, and the optimum interpolation weights used to form the assimilation increments at the analysis grid points.

4.3 Interpolation to model grid points

The analysis grid does not coincide with the model grid. This is to avoid redundant calculation near the Poles where Fourier truncation removes unstable modes from the model fields, and to introduce a degree of smoothing into the field of assimilation increments. Annex 4 gives details of the two grids. The increments on the model grid are calculated from those on the analysis grid simply by bi-linear interpolation in the horizontal and linear interpolation in log (sigma) in the vertical.

4.4 Specification of relaxation coefficients

Before the increments are added to the model values they are scaled by the relaxation coefficients λ in equation (4.1) which may vary with time, location, height and model variable. The choice of suitable values is largely the result of experimentation and experience, but the basic principles determining their specification have been outlined by Lyne (1979). Too large a value can lead to the generation of large amplitude gravity wave oscillations whilst too small a value results in data not being adequately assimilated. Their optimum values depend on the scale of the feature being analysed and the variable being observed. Current values of λ are given in Annex 10.

4.5 Geostrophic wind increments

During the early development of this assimilation scheme it was found that mass data were not always adequately assimilated when there were no supporting wind observations. This could be explained in terms of geostrophic adjustment theory which states that for most synoptic scales the mass field tends to adjust to the wind field when they are initially unbalanced (eg see Daley 1980). This was particularly true for surface pressure data. To overcome this problem geostrophic wind increments are calculated from the assimilation increments of the mass field and, after being suitably scaled, are added to the model wind fields after each full advection step. The scaling factor in equation (4.1) has a latitudinal dependence falling to zero near the equator, and current values are given in Annex 10. Similar techniques have been employed by other workers e.g. Kistler and McPherson (1975), Hayden (1973).

The geostrophic equations in σ co-ordinates are

$$\left.\begin{aligned} u &= \frac{-1}{af}\left(\frac{\partial \Phi}{\partial \phi} + C_p \sigma^K \theta \frac{\partial \pi}{\partial \phi}\right) \\ v &= \frac{1}{af\cos\phi}\left(\frac{\partial \Phi}{\partial \lambda} + C_p \sigma^K \theta \frac{\partial \pi}{\partial \lambda}\right) \end{aligned}\right\} \qquad (4.2)$$

where u and v are the westerly and southerly components of wind, f the Coriolis parameter, Φ the geopotential, θ potential temperature, and π is $(P_S/1000)^K$, P_S being the surface pressure. The constant K is the ratio of the gas constant R to the specific heat of the atmosphere at constant pressure C_p, a is the radius of the Earth, and ϕ and λ are latitude and longitude respectively. The geopotential Φ is found by integrating the hydrostatic equation so that at level k

$$\Phi_k = \int_{\sigma_k}^{1} R \pi \theta \sigma^{K-1} d\sigma + \Phi_s \qquad (4.3)$$

To calculate geostrophic wind increments the corresponding increment equations can be derived for each level.

$$\left.\begin{aligned}\Delta u_k &= \frac{-1}{af}\left(\frac{\partial\Delta\Phi_k}{\partial\phi} + C_p\sigma_k^K\Delta\theta_k\frac{\partial\Pi}{\partial\phi} + C_p\sigma_k^K\theta_k\frac{\partial\Delta\Pi}{\partial\phi}\right)\\ \Delta v_k &= \frac{1}{af\cos\phi}\left(\frac{\partial\Delta\Phi_k}{\partial\lambda} + C_p\sigma_k^K\Delta\theta_k\frac{\partial\Pi}{\partial\lambda} + C_p\sigma_k^K\theta_k\frac{\partial\Delta\Pi}{\partial\lambda}\right)\end{aligned}\right\} \quad (4.4)$$

$$\Delta\Phi_k = \int_{\sigma_k}^{1} R\left(\theta_k\Delta\Pi + \Pi\Delta\theta_k\right)\sigma^{K-1}d\sigma$$

where terms of $O(\Delta^2)$ have been neglected. The finite difference approximationsto these equations are given in Annex 5.

4.6 Hydrostatic temperature increments

To improve further the fitting of surface pressure data, temperature increments are calculated by a formula derived from the hydrostatic equation. In pressure co-ordinates equation (4.3) becomes

$$\Phi_k = \int_{P_k}^{P_s} R\,\theta\left(\frac{P}{1000}\right)^K \frac{dp}{p} + \Phi_s$$

and with the addition of surface pressure and temperature increments

$$\Phi_k = \int_{P_k}^{P_s+\Delta P_s} R(\theta+\Delta\theta)\left(\frac{P}{1000}\right)^K \frac{dp}{p} + \Phi_s$$

The hydrostatic temperature increments are defined as those which lead to no change in the value of Φ_k when the surface pressure is incremented by ΔP_s.

$$\int_{P_k}^{P_s+\Delta P_s} R(\theta+\Delta\theta)\left(\frac{P}{1000}\right)^K \frac{dp}{p} = \int_{P_k}^{P_s} R\theta\left(\frac{P}{1000}\right)^K \frac{dp}{p}$$

If $O(\Delta^2)$ terms are neglected this becomes

$$\frac{\Delta p_s}{p_s}\,R\,\theta_s\left(\frac{p_s}{1000}\right)^K = -\int_{p_k}^{p_s} R\,\Delta\theta\left(\frac{p}{1000}\right)^K \frac{dp}{p}$$

It is now assumed that $\Delta\theta$ is independent of the pressure p and that the level k is removed to $P_k = 0$. Then the hydrostatic temperature increments are given by

$$\Delta\theta = -K\,\theta_s\,\frac{\Delta P_s}{P_s} \qquad (4.5)$$

$$\frac{\partial D}{\partial t} = \frac{1}{a\cos\phi}\left[\frac{\partial A_\lambda}{\partial\lambda} + \frac{\partial}{\partial\phi}(A_\phi\cos\phi)\right] + \mu\nabla^2 D$$

$$\frac{\partial\zeta}{\partial t} = \frac{1}{a\cos\phi}\left[\frac{\partial A_\phi}{\partial\lambda} - \frac{\partial}{\partial\phi}(A_\lambda\cos\phi)\right] \qquad (4.7)$$

where ζ is the vorticity. Hence the divergence diffusion has no effect on vorticity whilst damping the divergence.

The finite difference approximations for the damping terms are given in Annex 5, and the current value of μ in Annex 10. This value is restricted by the stability criterion which may be shown to be

$$\mu \leq \Delta t\left(\frac{\Delta s^2}{\Delta t^2} - c^2\right)$$

Where Δt is the time step of each adjustment step, Δs is the grid length, and c is the phase speed of the fastest move gravity wave. The damping is also applied during the six hour forecast which produces the background field for the next analysis stage (see section 3).

References

Birch, N T & Lyne, W H 1980 The FGGE data assimilation scheme. Unpublished M.O. internal note Met O 20 Tech. Note II/142.

Daley, R 1980 On the optimal specification of the initial state for deterministic forecasting Mon. Wea. Rev., 108, 1719-1735.

Davies, H C & Turner, R E 1977 Updating prediction models by dynamical relaxation: an examination of the technique. Quart. J. R. Met. Soc., 103, 225-245.

Dey, C H 1978 Noise suppression in a primitive equation model. Mon. Wea. Rev., 106, 159-173.

Gandin, L S 1963 Objective analysis of meteorological fields. Pub. Izdatel, Leningrad. Translated by Israel program for scientific translations. Jerusalem 1968.

Ghil, M, Halem, M & Atlas, R 1979 Time-continuous assimilation of remote-sounding data and its effect on weather forecasting. Mon. Wea. Rev., 107, 140-171.

Hayden, C M 1973 Experiments in the four-dimensional assimilation of Nimbus 4 SIRS data. J. Appl. Met., 12, 425-436.

Lyne, W H, Swinbank, R & Birch, N T — 1982 — A data assimilation experiment and the global circulation during the FGGE special observing periods. Quart. J. R. Met. Soc., 108, 575-594.

Miyakoda, K, Stickler, R F & Chludzinski, J — 1978 — Initialization with the data assimilation method. Tellus, 30, 32-54.

Rutherford, I D — 1972 — Data assimilation by statistical interpolation of forecast error fields. J. Atmos. Sci., 29, 809-813.

Temperton, C — 1983 — To be published.

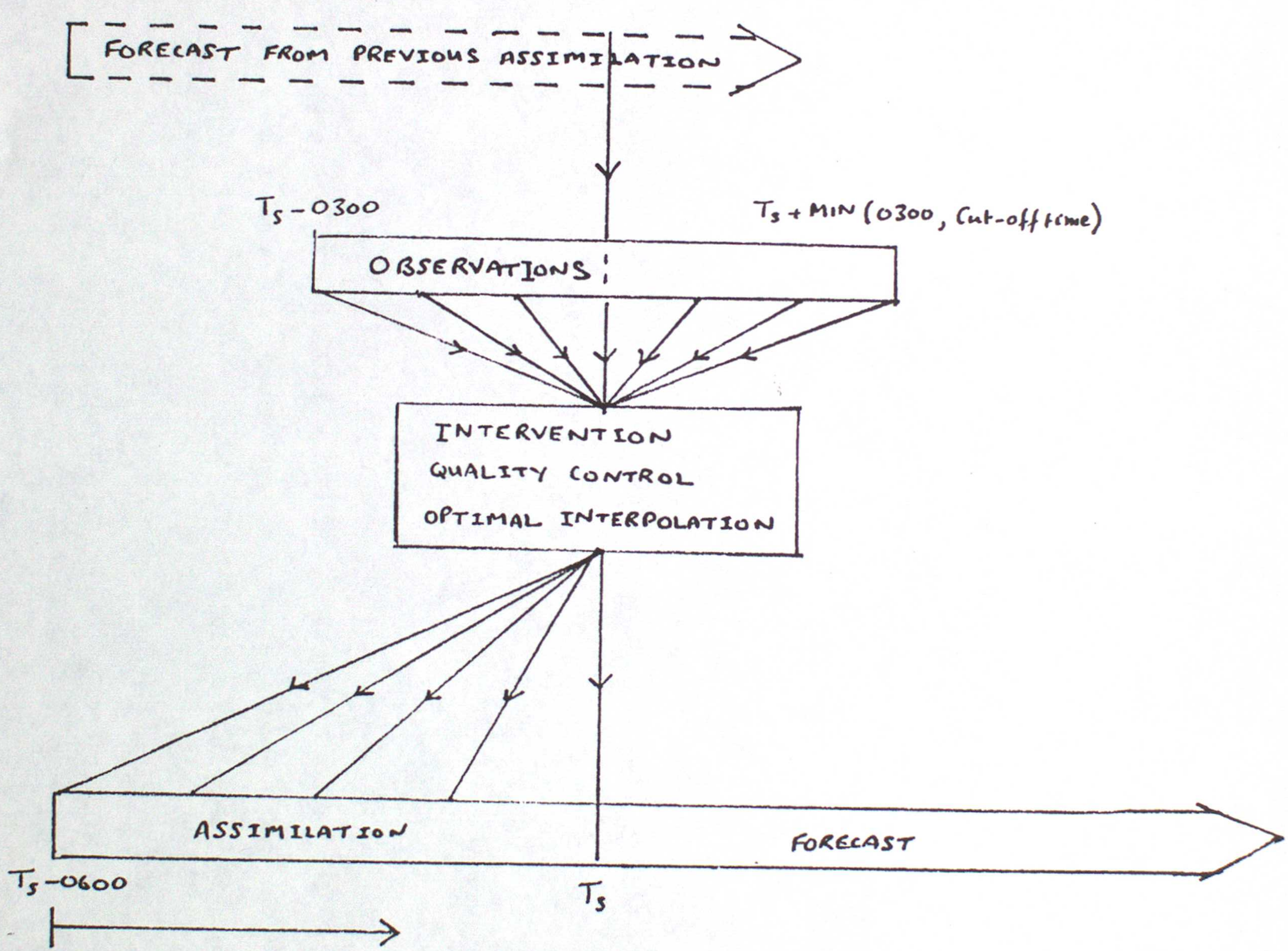


Figure 2.1 The assimilation cycle.

Annex 1 Design of the analysis program

1. General

The analysis program was designed to be flexible, easily modified and, without significant loss of efficiency, to be used both operationally and as a research tool. This was achieved by:

i. The adoption of OLYMPUS conventions and structure,

ii. The ability to handle a variety of data grids,

iii. Common code for both the quality control stages (Modes 1 and 2) and the calculation of optimum interpolation weights (Mode 3).

iv. Overlapping of I/O with CPU processing.

v. Packed format for the data files to save I/O time (see Annex 3).

It was further intended to use standard FORTRAN wherever possible, but the requirement for operational efficiency and inadequacies of the early compilers have resulted in extensive use of non-standard vector FORTRAN and special 'Q8' calls. The remaining code, including the control logic, is basically ANSI 1966 standard, but features of ANSI 1977 FORTRAN have been introduced as they have become available in successive releases of CYBER 205 FORTRAN compilers. CYBER 205 extensions to standard FORTRAN have been used, in particular BIT and DESCRIPTOR variables for data processing (but not program control), and also dynamic space allocation accessed via DESCRIPTORS.

2. The OLYMPUS system

The adoption of the OLYMPUS system on a vector processor has necessitated the extension of its conventions. These are detailed below together with a brief description of the existing conventions.

ble A1.1 Analysis supplied OLYMPUS routines

Routine	Number	Description
LABRUN	1.1	Label this run (not used)
CLEAR	1.2	Clear all variables and arrays to machine zero
PRESET	1.3	Set default values
DATA	1.4	Read changes to defaults from input
AUXVAL	1.5	Calculate derived values
INITIAL	1.6	Initial calculations (Open data sets, read header records, extract default values pertaining to datasets and their derived values, and initialize buffers in blank common)
START	1.8	Start the calculation
OUTPUT(1)	3.1	Initial output (not used)
STEPON	2.1	Advance the calculation (step through rows of data. If Mode 1 or 2 do data checking, if Mode 3 do calculation of optimum interpolation weights)
OUTPUT(2)	3.1	Printout from this step
TESEND	4.1	Test whether required number of steps completed
OUTPUT(3)	3.1	Final output (not used)
ENDRUN	4.2	Close datasets, print message and stop. Cause dump if error messages printed.

The initial letters of variable names are used to distinguish where it is assigned as well as its type. The standard conventions are given in table A1.2.

Table A1.2 Standard OLYMPUS coding conventions

Description	Type		
	REAL	INTEGER	LOGICAL
Variables held in COMMON	A-H,O, Q-Y	L-N	LL-NL
Subroutine dummy arguments	P	K	KL
Variables local to routine or program unit	Z	I	IL
Loop variables	-	J	-

These do not contravene the default implicit typing of variables by FORTRAN. The extension for the non-standard BIT and DESCRIPTOR variables available on the CYBER 205 is analogous to the OLYMPUS convention for LOGICAL variables. BIT variables have a B as the second letter and descriptor variables have D as the second (or third for BIT DESCRIPTORS) and last letter. The latter refinement was included so as not to conflict with a current Meteorological Office convention. These conventions are summarized in table A1.3 where it can be seen they do not contravene the standard OLYMPUS conventions. In practice most DESCRIPTOR variables are local to routines.

Table A1.3 Extensions to OLYMPUS coding conventions

Description	Type			
	BIT	DESCRIPTOR		
		REAL	INTEGER	BIT
Variables held in COMMON	AB-HB, OB, QB-YB	AD....D-HD....D OD....D, QD...D-YD...D	LD....D-ND...D	ABD...D-HBD....D, OBD....D QBD....D-YBD....D
Subroutine dummy arguments	PB	PD....D	KD....D	PBD....D
Local variables	ZB	ZD....D	ID....D	ZBD....D

Generally, significant variables are placed in COMMON rather than passed in argument lists. This is usually more efficient and leads to logically structured data.

3. STEPON

The main OLYMPUS loop is over observational data rows (or bands) with NBAND rows available for processing at each step and ND1 data bands altogether. The control of the I/O and of the calculation at each step (NSTEP) is performed from within the routine STEPON and is depicted in the flow charts figures A1.2 to A1.7. Brief descriptions of the routines are given in Table A1.4 but program documentation should be consulted for further details.

Table A1.4 Analysis subroutines called from STEPON

Routine	Number	Description
RDINFL	2.10	Reads INDEX and FLAG data sets using overlapped I/O
RDATA	2.11	Reads DATA, DOER and DPER using overlapped I/O
RFG	2.12	Reads BACK using overlapped I/O
RFE	2.13	Reads YFE5DEG using overlapped I/O
WDATA	2.14	Writes INCR,DOER,DPER and OBDATA using overlapped I/O
WWTS	2.15	Writes WEIGHTS using overlapped I/O
VSWAP	2.16	Swaps arrays and some buffers
VEXTRCT	2.2	Extracts details of observations from compact formats of INDEX and FLAG
GRID	2.22	Set up arrays of grid points for current data band. Points are observation points for quality control (Modes 1 and 2) and analysis grid points for optimal interpolation (Mode 3)
BLWTS(n)	2.3	Calculate bi-linear interpolation weights to interpolate background values (n = 1) or background error values (n = 2) to grid points
HINT(n)	2.4	Interpolate horizontally background values (n = 1) or background error values (n = 2) to grid point positions
VINT(n)	2.5	Interpolate vertically background values (n = 1) or background error values (n = 2) to grid point levels
VOBSERR	2.24	Calculate observation errors for all data in current band from internal tables
VFORM(n)	2.20	Format observation errors (n = 1), increments (n = 2) or background error values (n = 3) into observational data set format (DATA) and copy to appropriate buffers
VCHECK1	2.8	Flag data deviating by more than a certain number of standard deviations of expected error from background
STORE	2.9	Calculate certain values for current data band

Table A1.4 (Con't)

Routine	Number	Description
SELECT	-	Select nearest observations to a group of grid point positions
SELDAT	-	Select best data from nearest observations for a group of grid point variables, levels and positions
VBUILD	-	Construct the optimal interpolation equations for a group of grid point variables, levels and positions
VSOLVE	-	Solves optimal interpolation equations (called from VBUILD)
VCHECK2	-	Interpolate observations to a group of (observation) grid point variables, levels and positions, and flag data deviating by more than a certain number of standard deviations of expected error from interpolated values.
VFORMWT(n)	2.26	Set up record header for current WEIGHTS record (n =-1), sort weights and pointers for group of grid point variables, levels and positions (n = 0), copy weights and pointers in packed format to output buffer and reset values in header record (n = +1)

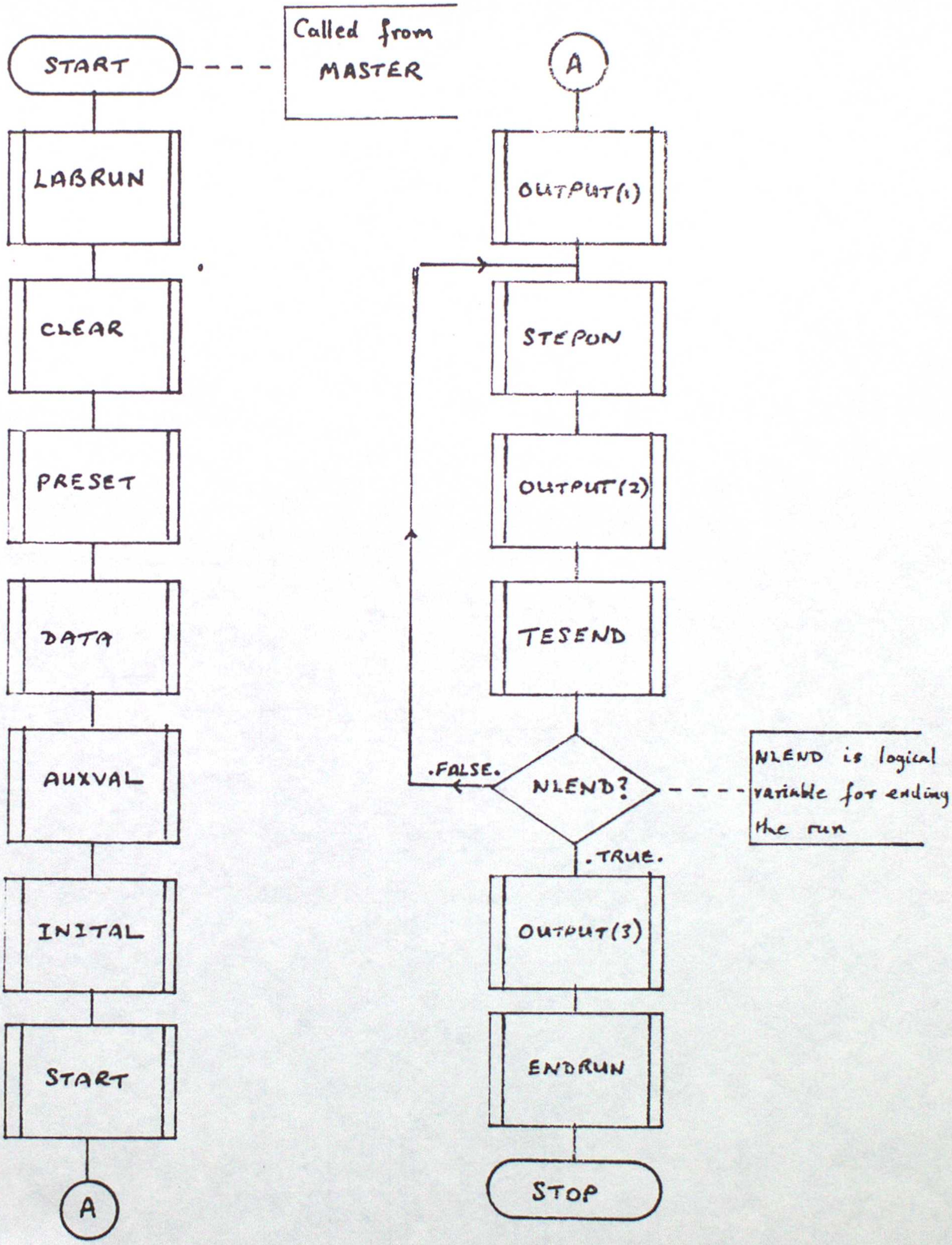


Figure A1.1 OLYMPUS routines supplied by analysis.

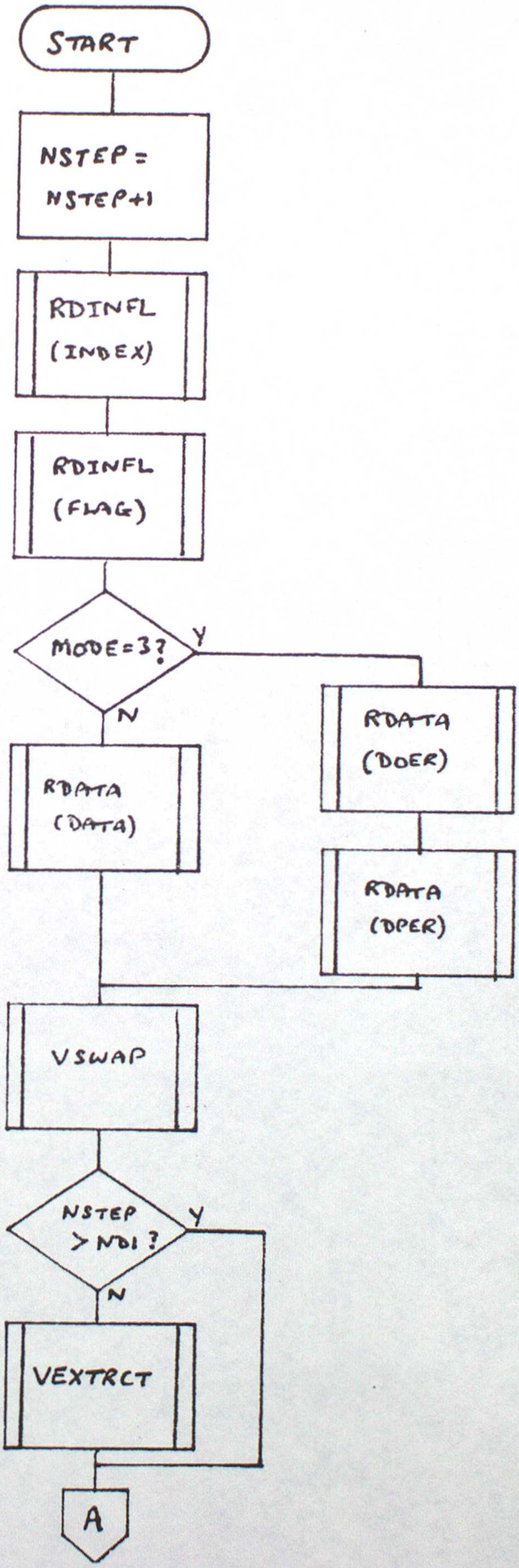


Figure A1.2 STEPON : input section.

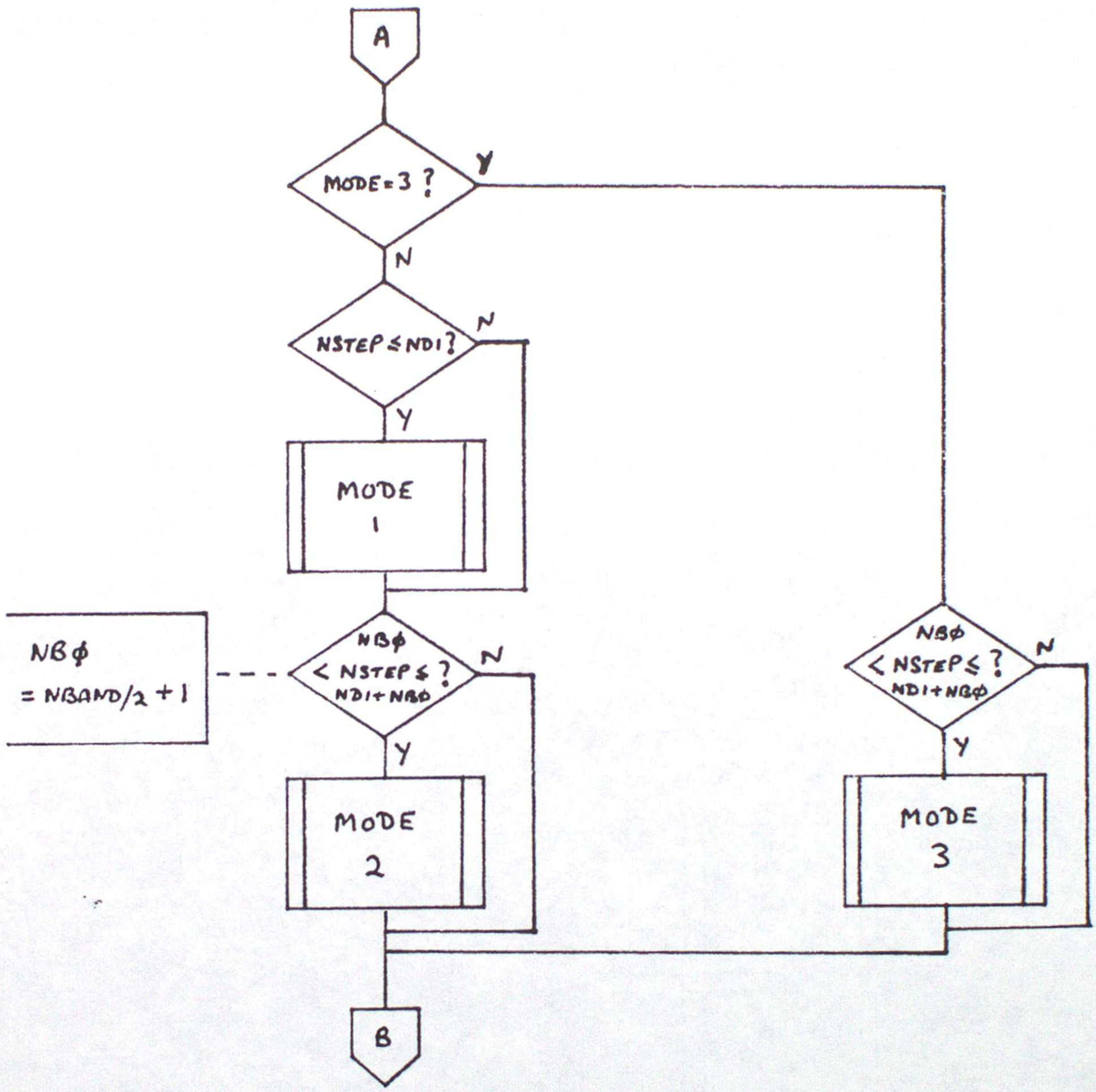


Figure A1.3 STEPON : processing section.

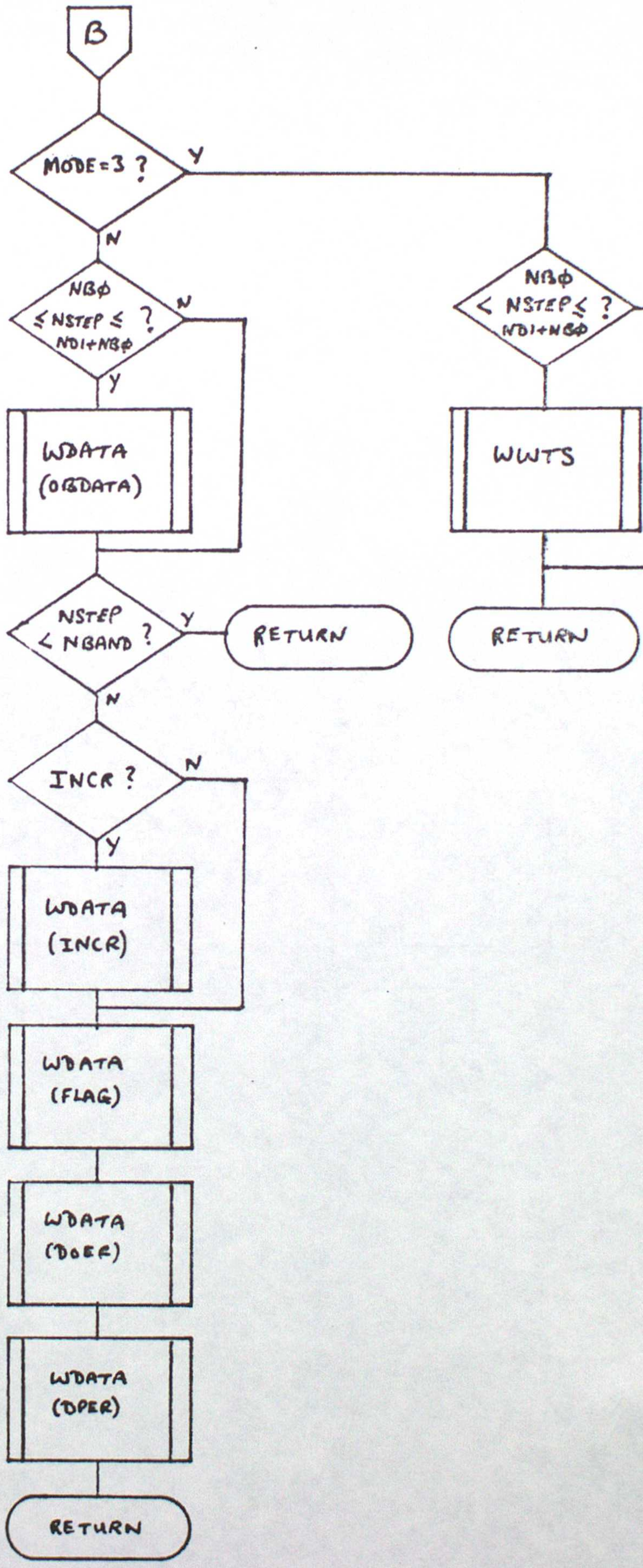


Figure A1.4 STEPON : output section.

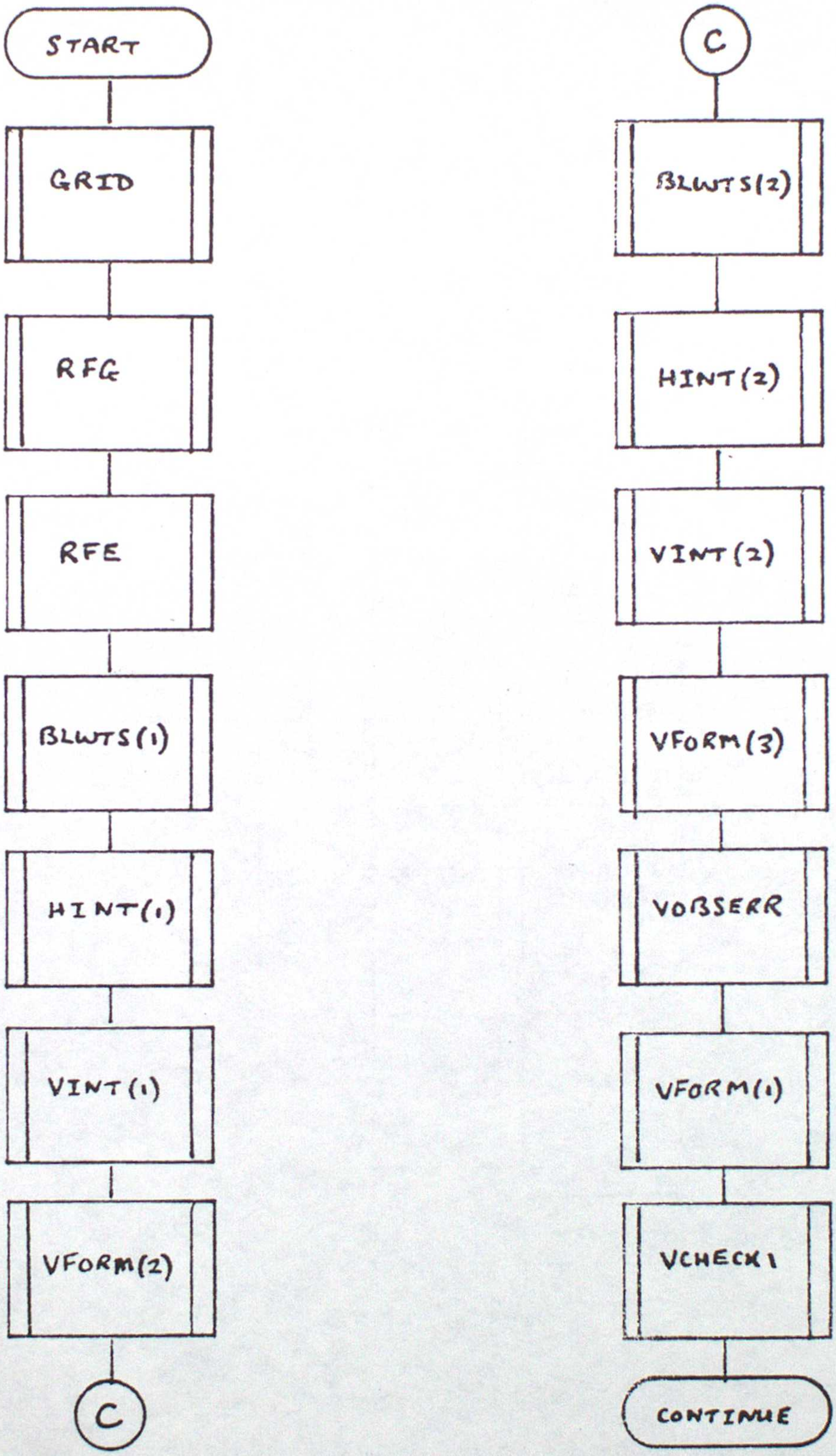


Figure A1.5 STEPON : MODE 1

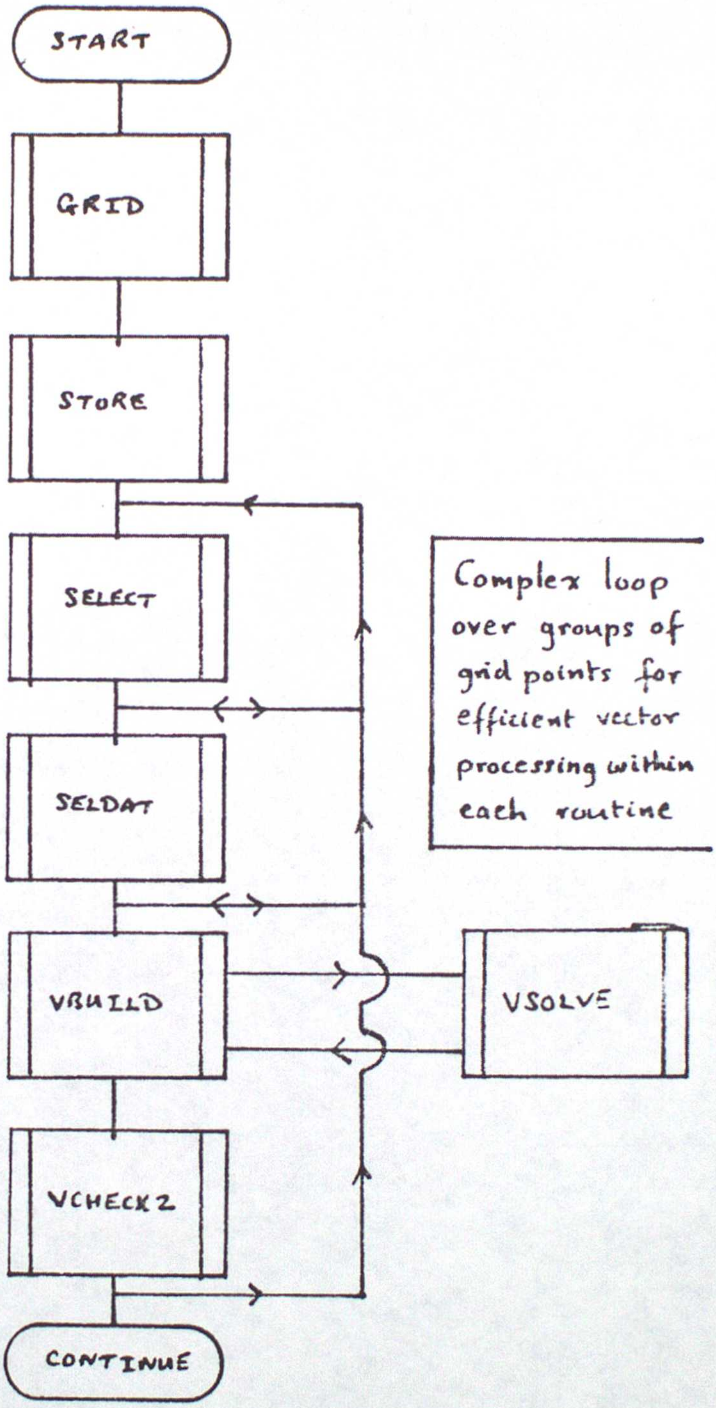


Figure A1.6 STEPON : MODE 2

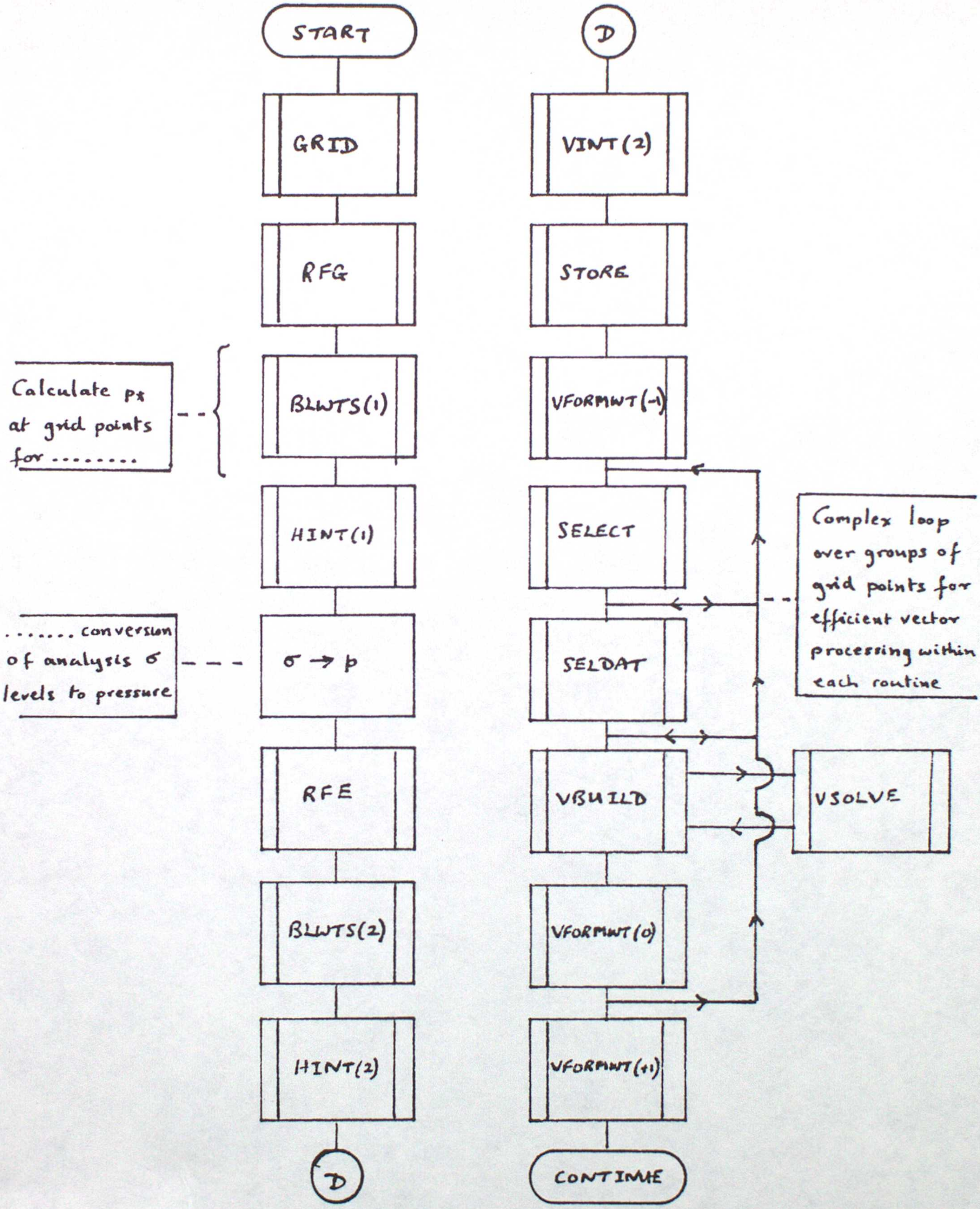


Figure A1.7 STEPON : MODE 3

Annex 2 Design of the assimilation program

The code for assimilating observations into the model was designed as a separate package to be inserted into a version of the forecast model which performs the calculations on a row by row basis. This method of processing was adopted because there was insufficient computer storage for the model, observations and weights when organised on the fields basis of the standard forecast model (see FC).

The flow diagrams in figures A2.1 to A2.3 show the basic structure of the program and brief descriptions of the subroutines are given in Tables A2.1 to A2.3. All the preliminary processing is performed in the CONTROL section where the entire observations file is read and stored for future use. The subroutine CYCLE is then called once for each timestep and from within this routine all the calls to the assimilation, physics and dynamics routines are made for each row of the model, cycling from north to south. The I/O calls to read the model fields at timestep (n - 1), read the weights data and write the model fields at timestep (n), are also performed within subroutine CYCLE. All these calls are staggered to allow the appropriate number of rows to be available for each routine at the same time level.

Table A2.1 Assimilation subroutines called from CONTROL

Routine	Description
HKCHECK	Checks start time is consistent with background data/time
CONSTS	Calculates numerous constants to be used by subsequent routines
CYCLE	Secondary control routine
INCTIM	Calculates date/time of write-up

Table A2.2 Assimilation subroutines called from CYCLE

Routine	Description
PHYSFIX	Called on first timestep only, zeroes rain, resets boundary depth and surface land temperatures
BLINT	Bi-linear horizontal inerpolation of model fields to observation points and linear interpolation in log sigma to model levels. Calculates observation - model differences at observational data location.
OIINT	Interpolates observation - model differences to analysis grid points using optimal interpolation weights to give assimilation increments
GRIDCHZ	Vertical interpolation of assimilation increments from analysis grid to model grid
GRIDCHX	Horizontal interpolation of increments from analysis grid to model grid along a line of latitude
GRIDCHY	Horizontal interpolation of increments from analysis grid to model grid along a line of longitude.
GSTPRC	Calculates geostrophic wind increments
PHYSICS	Control routine for physics package including reformatting from rows to fields
POLAR	Performs averaging for polar rows
ASMDTA	Adds scaled assimilation increments to model fields and calculates and adds hydrostatic temperature increments
PTRADJ	Adjustment step for surface pressure, potential temperature and humidity. Calculate geopotential and vertical velocity
UVADJ	Adjustment step for velocity components
TRADVI	First Lax-Wendroff step for potential temperature and humidity advection
UVADVI	First Lax-Wendroff step for velocity component advection
TRFILT	Diffusion for potential temperature and humidity
UVFILT	Diffusion for velocity components

Table A2.2 (Con't)

Routine	Description
TRADV2	Second Lax-Wendroff step for potential temperature and humidity advection
UVADV2	Second Lax-Wendroff step for velocity component advection
ASMDTAG	Adds geostrophic wind increments to model field

Table A2.3 Assimilation subroutines called from PHYSICS

Routine	Description
BNDARY	Boundary layer parametrization
HELIOS	Radiation routine
PHYSFILT	Filter of physics increments
RAIN	Dynamic rain routine
SHOWER	Convective rain routine

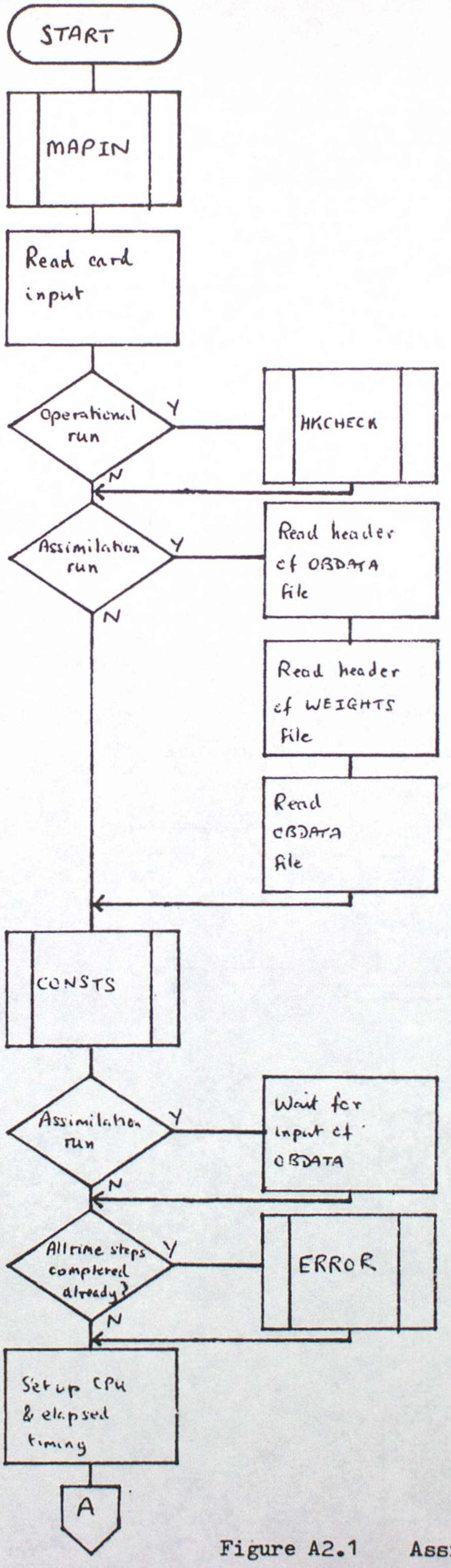


Figure A2.1 Assimilation program CONTROL.

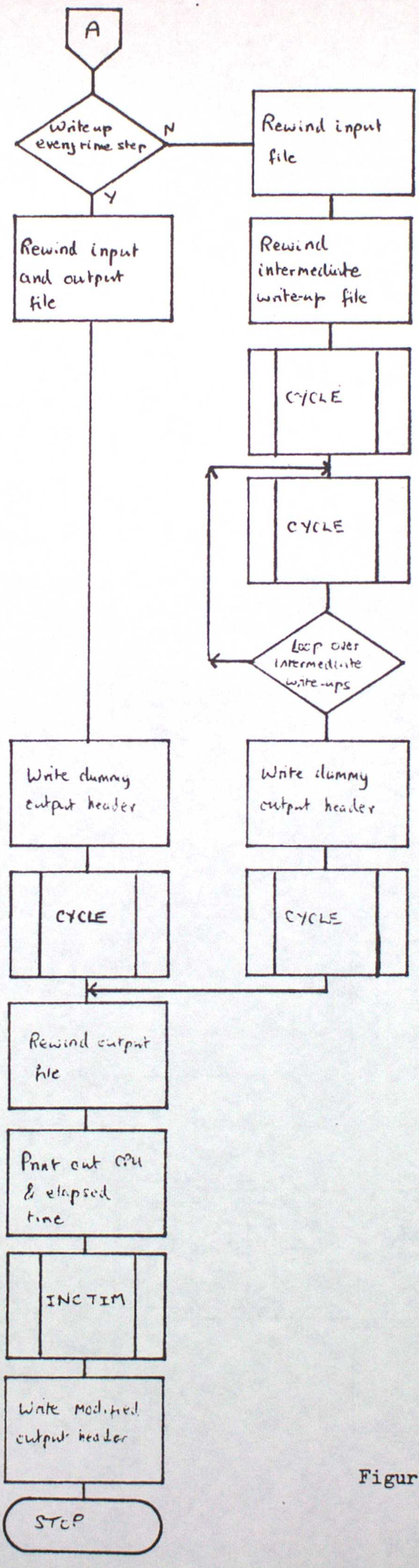


Figure A2.1 (cont)

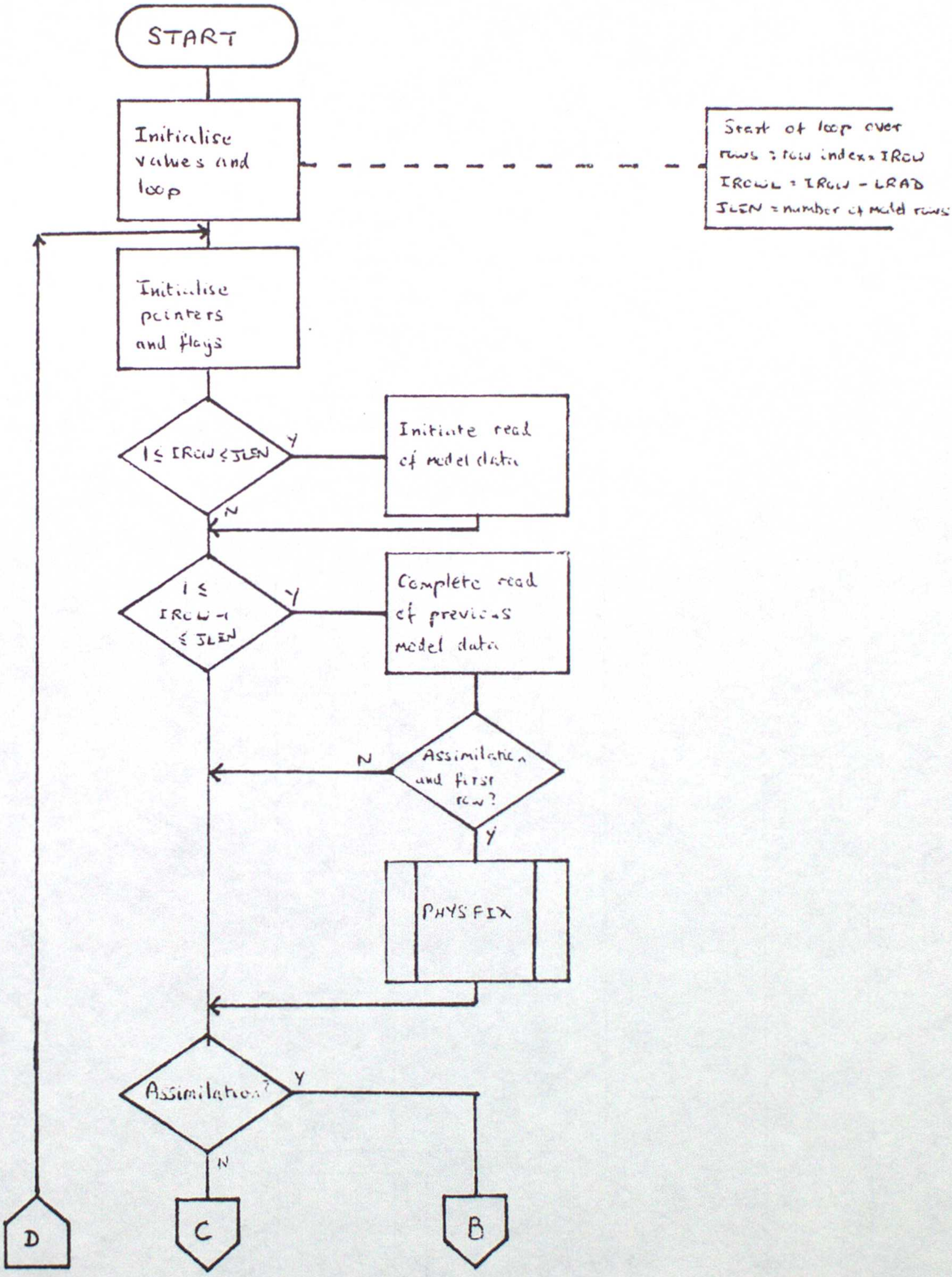


Figure A2.2 Assimilation subroutine CYCLE.

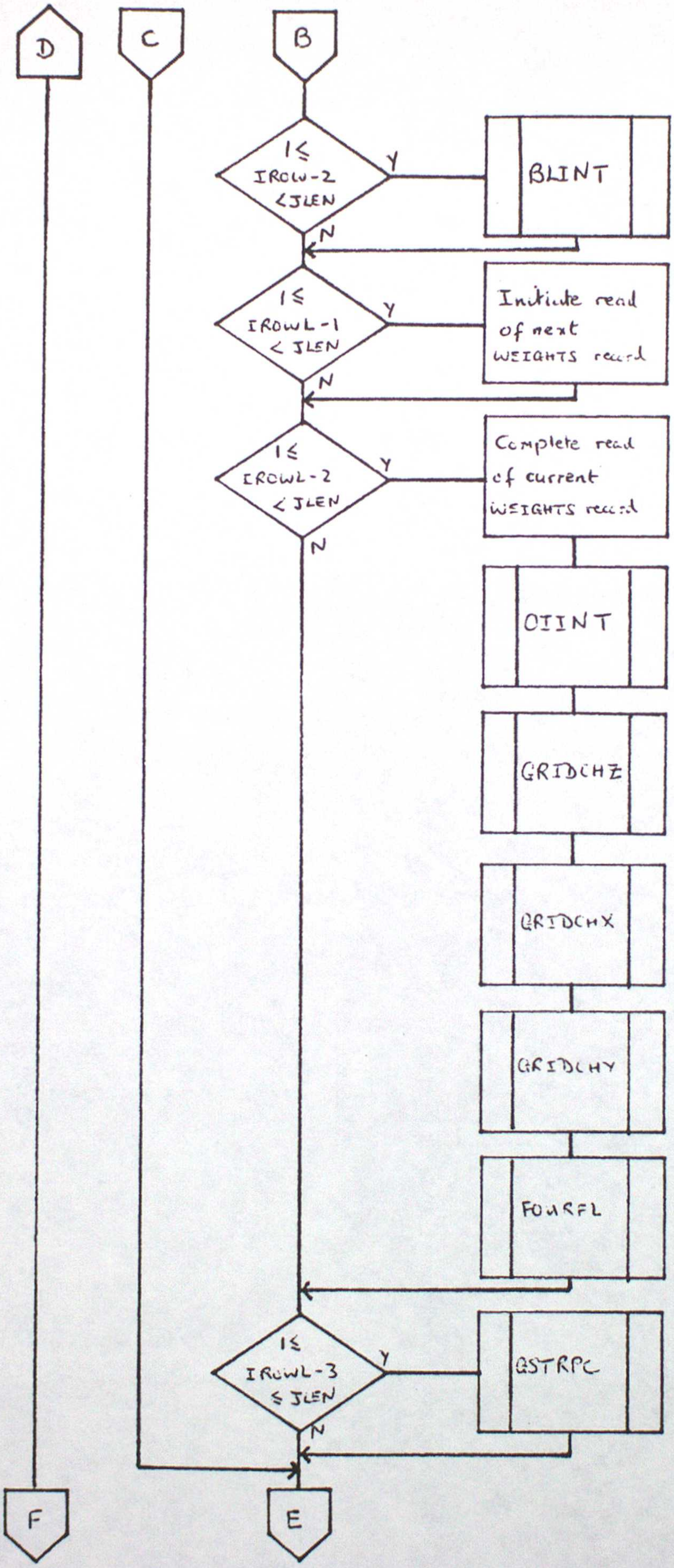


Figure A2.2 (cont.)

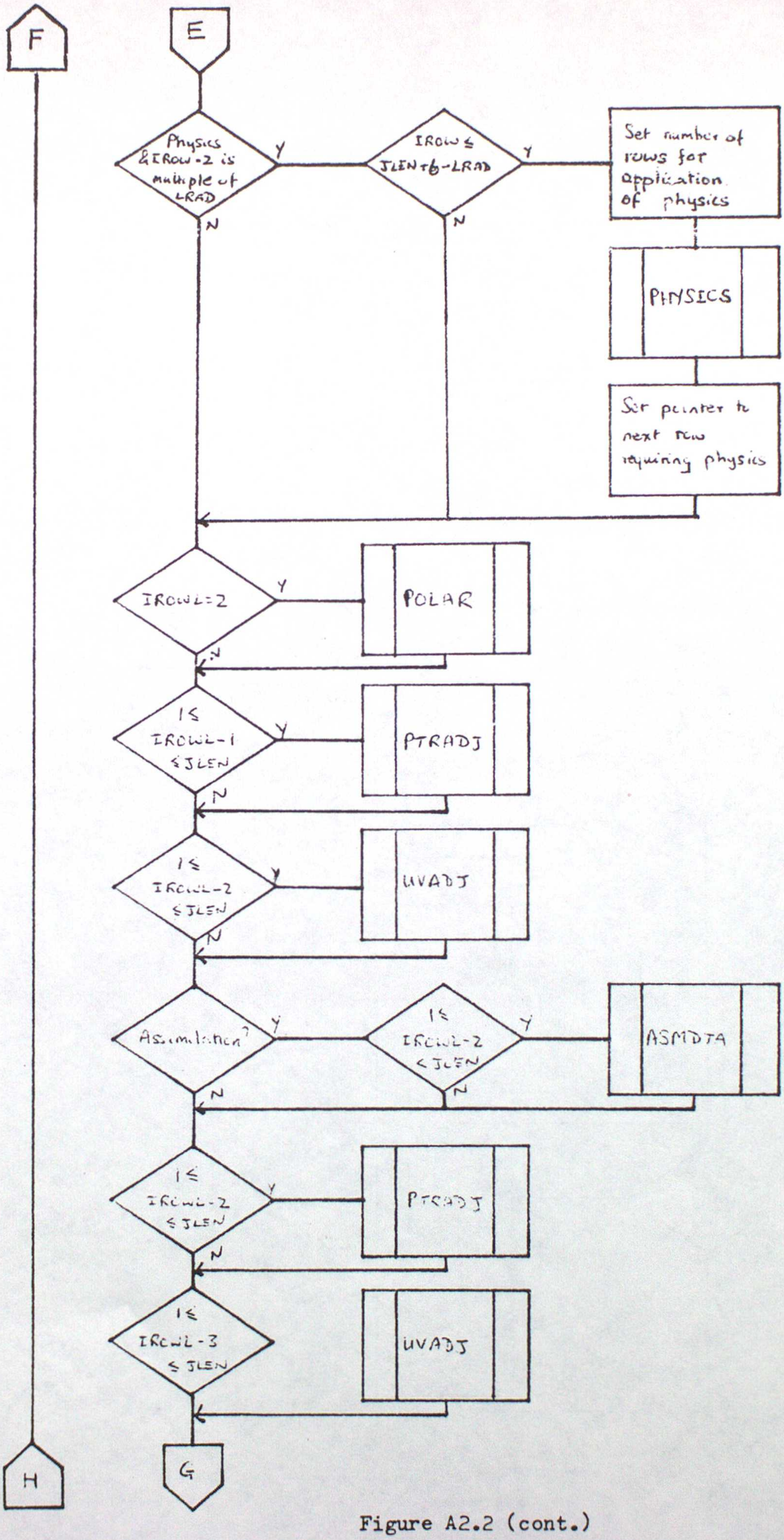


Figure A2.2 (cont.)

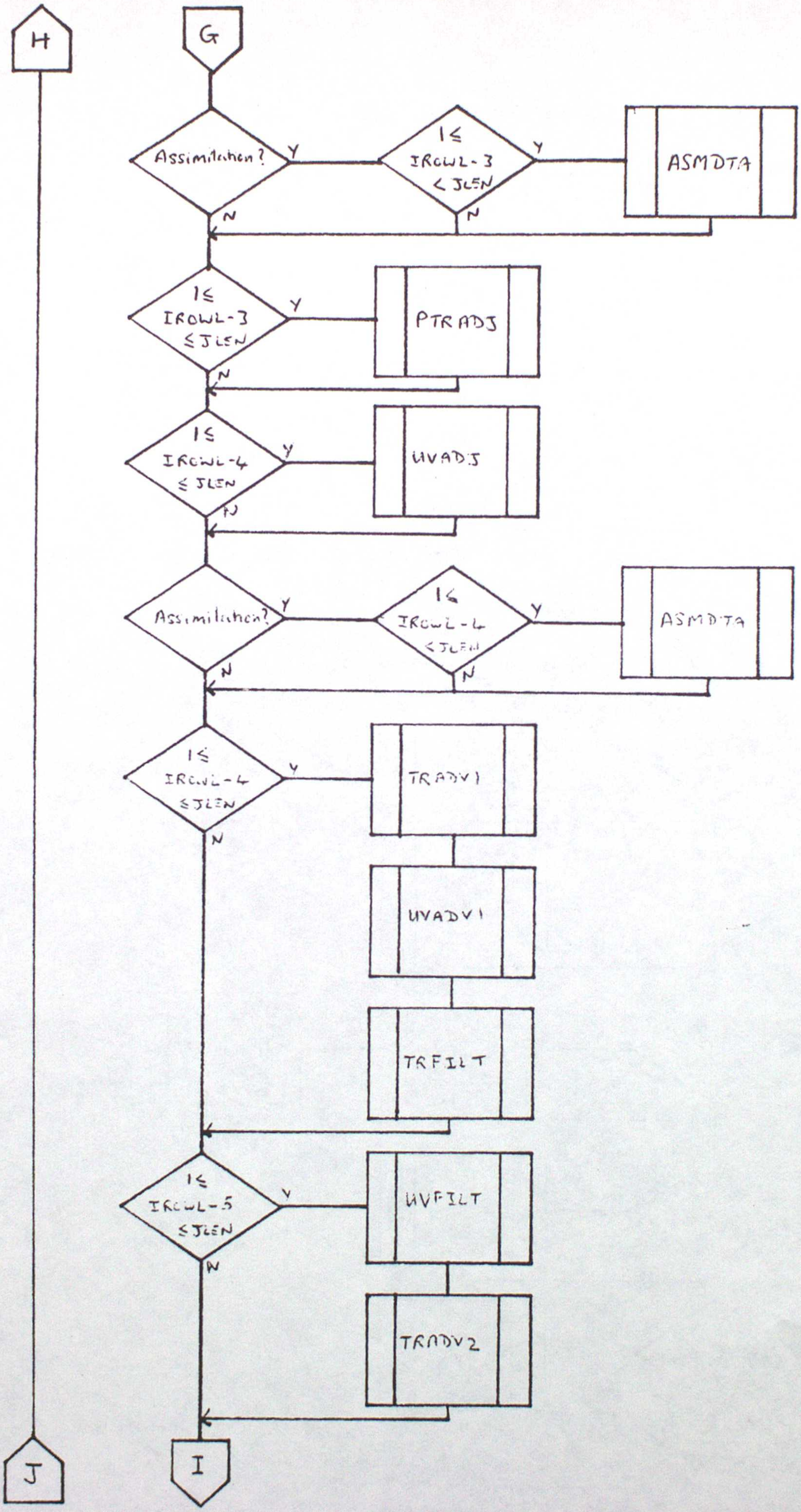


Figure A2.2 (cont.)

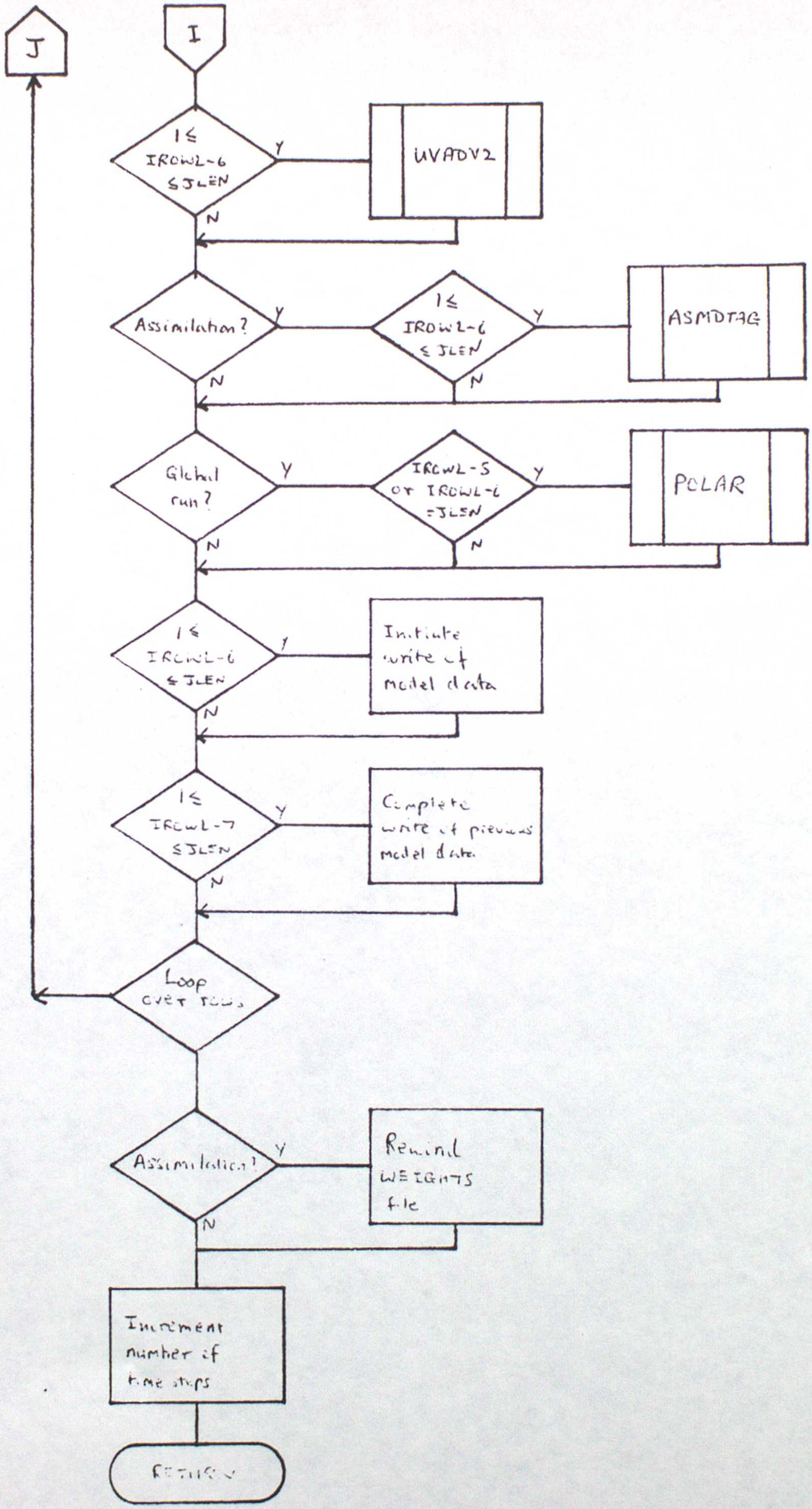


Figure A2.2 (cont.)

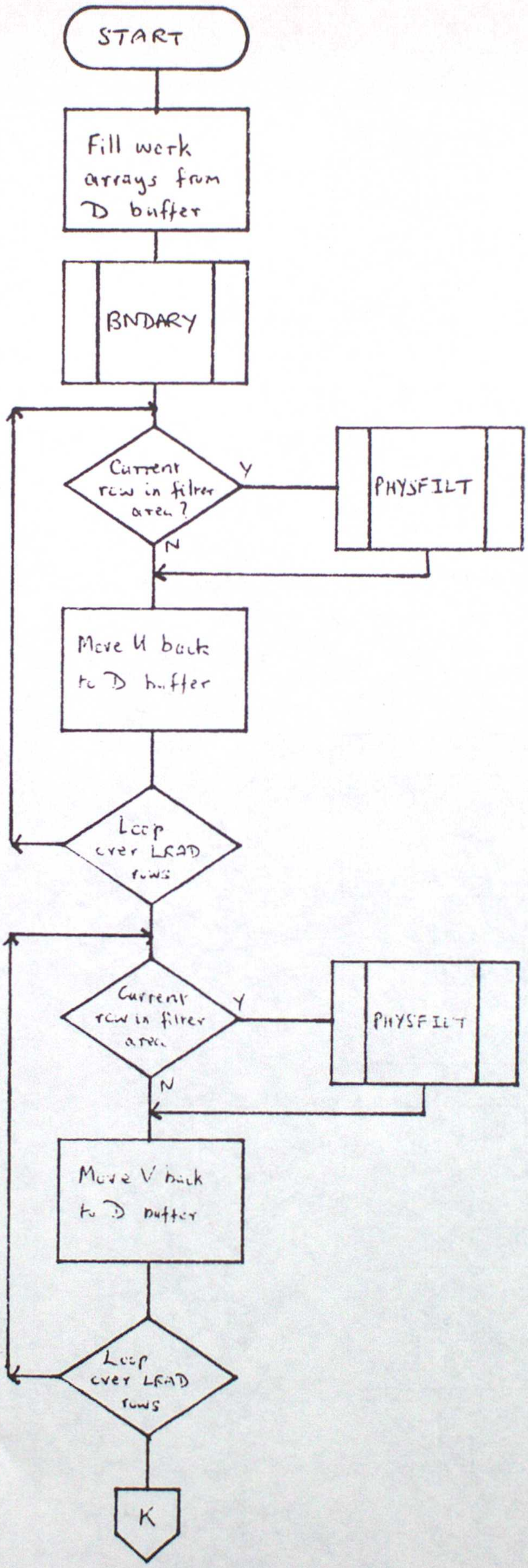


Figure A2.3 Assimilation subroutine PHYSICS

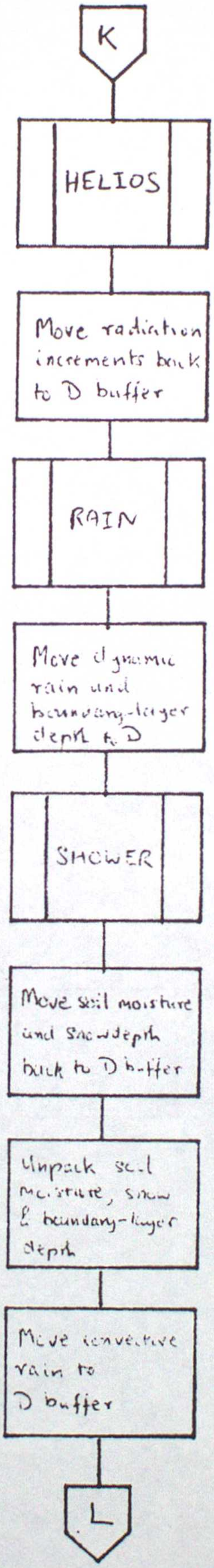


Figure A2.3 (cont.)

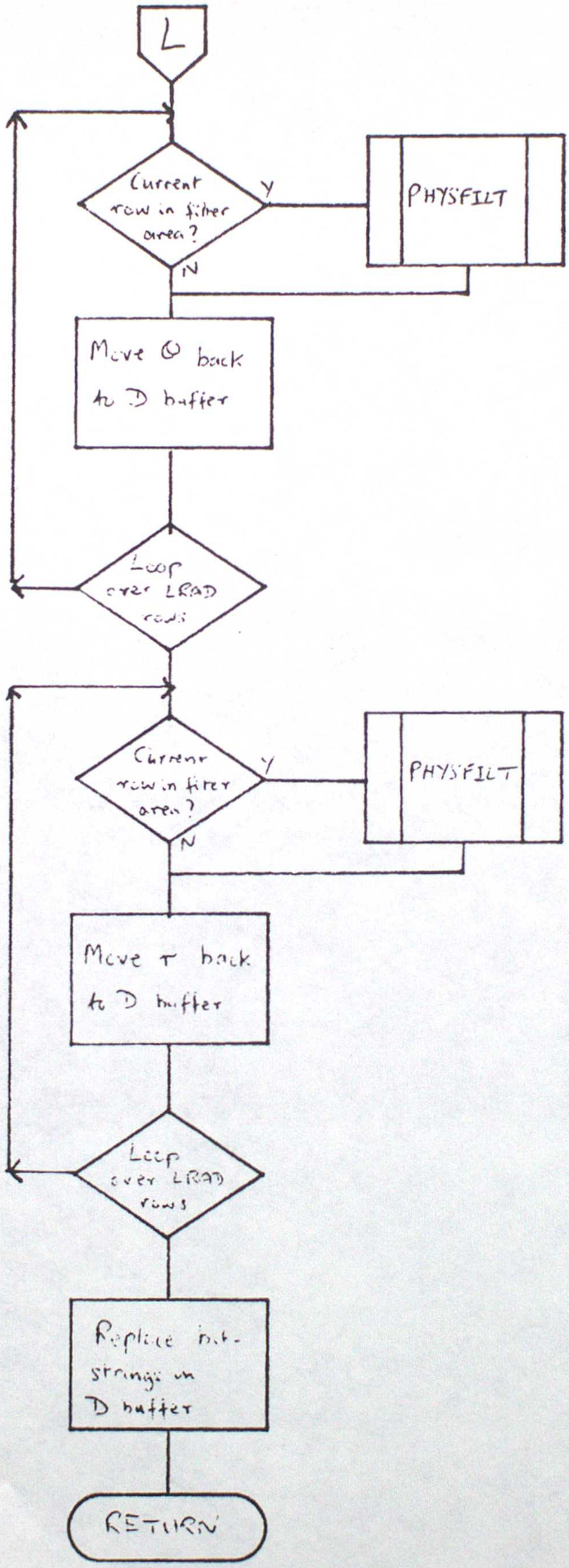


Figure A2.3 (cont.)

Annex 3 Data Files

1. Description

a. Files input to optimum interpolation stage

Name	Contents
INDEX	Positional parameters of observational data
FLAG	Data indicators, quality control flags and pressure levels of observational data
DATA	Observational data
BACK	Model background or first guess
YFE5DEG	Model errors on coarse (5 degree) grid

b. Files output from optimum interpolation stage

Name	Contents
DOER	Observational errors at data locations
DPER	Model errors at data locations
INCR	Differences between observations and model values at data locations
FLAG2	As FLAG but with quality control indicators set
OBDATA	File formed by merging INDEX, FLAG and DATA together with the weights and indices necessary to interpolate model fields to observation points
WEIGHTS	Optimum interpolation weights together with indices to locate relevant data and grid points

c. Files input to the assimilation stage (excluding forecast files)

OBDATA)
) As output from optimum interpolation
WEIGHTS)

A flow diagram charting the interaction of the files with the analysis and assimilation files is given in figure A3.1.

2. Formats

All files have undefined record type (RT=U) on the CYBER, and where possible indications have been made as to how the files may be supported on the IBM.

a. Observational data files INDEX, FLAG and DATA

i. Header records

INDEX, FLAG and DATA share a common header record format which was originally advocated to be adopted by all data sets. Because of this many elements are not applicable to observational data, but are nonetheless described here to emphasis the format's generality. The application of this header format to the other data files is still intended in the future.

Each header record consists of five sections:

Fixed length header

Record length header

Integer header

Real header

Character header

These are contained in the first 2 sectors (1024 words) of the relevant CYBER data set, or in the first 5 logical records of an IBM data set.

The fixed length header contains general information concerning the contents of the data set and pointers to locate the sections. The record length header is essential to enable the variable length records of these data sets to be read by the concurrent I/O subroutines of the CYBER. The remaining headers contain information necessary to specify the structure of the data e.g. grid definitions, or to describe the data itself, e.g. plain language message in character header. Information which is data dependent e.g. number of observations in data band, are not held in the header, but within the records themselves.

In the following, each element of the header is a 64 bit word on the CYBER, a 32 bit word on the IBM, and on the CYBER the headers follow immediately one after the other.

Fixed length header

20 words long, all integers: Identifier I

Word	Description
I(1)	Max. length of subsequent logical data records in words
I(2)	Number of subsequent data records (Ø if not known)
I(3)	Ø for variable length records, 1 for fixed length records
I(4)	Type of data Ø mixed 1 integer 2 real 3 character 4 half word integers (not supported on CYBER) 5 half word reals (not supported on IBM) 6 binary (not supported on IBM)
I(5)	Year (e.g. 1979)) Data time
I(6)	Month (1 to 12)) Data time
I(7)	Day (1 to 31)) Data time
I(8)	Hour (0 to 23)) Data time
I(9)	Minute (0 to 59)) Data time
I(10)	Year) Validity time
I(11)	Month) Validity time
I(12)	Day) Validity time
I(13)	Hour) Validity time
I(14)	Minute) Validity time
I(15)	Length of record length header (Ø if none)
I(16)	Length of integer header (Ø if none)
I(17)	Length of real header (Ø if none)
I(18)	Length of character header (Ø if none)
I(19)	) Spare
I(20)	) Spare

Record length header

I(2) words long: Identifier L

L(1) to L(I(2)) to contain the length of subsequent records in words

Integer header

M(2) + 8 words long: Identifier M

Word	Mnemonic	Description
M(1)	LTYPE	Grid type, 1 for lat/long, 2 Gaussian, 3 quasi-homogeneous, 4 for 2 and 3 combined, Ø reserved for observational data.
M(2)	N1	No of latitude rows
M(3)	N2	No of longitude points in each row. (Minimum number for LTYPES 3 and 4, Ø for observational data)
M(4)	N3	-1, 0, +1 for stagger of LTYPES 3 and 4, Ø otherwise
M(5)	NVAR	No. of multi-level variables
M(6)	NSVAR	No. of surface variables
M(7)	NLEV	No. of vertical levels excluding surface, Ø for observational data
M(8)	NLEVT	Vertical level type, 1 for pressure, 2 for sigma levels
M(9) to M(8 +N1)	NLON(J)	No. of longitude points in each row. Not applicable for observational data.

Real header

5 + 2(N1+NVAR+NSVAR)+NLEV words long: Identifier R

Word	Mnemonic	Description
R(1)	R1	First latitude of grid
R(2)	R2	Last latitude of grid
R(3)	R3	First longitude of grid
R(4)	R4	Last longitude of grid

Word	Mnemonic	Description
R(5)	R5	Not used at present
R(6) to R(5 + N1)	ALAT(J)	Latitudes of each row of grid points
R(6 + N1) to R(5 + 2 N1)	FLON(J)	Starting longitude of each row of grid
R(6 + 2 N1) to R(5 + 2 N1 + +NVAR+NSVAR)	SLAT(J)	Stagger, in terms of grid length, of variables in meridional direction with respect to grid
R(6 + 2 N1 + +NVAR+NSVAR) to R(5 + 2(N1+NVAR+ NSVAR))	SLON(J)	Stagger, in terms of grid length, in zonal direction
R(6 + 2(N1+NVAR+ NSVAR)) to R(5 + 2 (N1+NVAR+NSVAR)+ NLEV)	ALEV(J)	Vertical levels at which data occurs excluding surface

SLAT, SLON and ALEV are not applicable to observational data.

Character header

No specification at present.

ii. Data records

The contents of these records have been arranged as vectors, and a header section is incorporated at the start of each record to facilitate access of the vectors. These vectors all start on 64 bit full word boundaries.

INDEX data records

The header section consists of 18 full word integers, the last 13 of which give the start locations of the constituent vectors. (These are the absolute locations of the first bit of each vector)

Word	Mnemonic	Description
1	NOBS	No. of observations in record ie the length of each constituent vector.
2	LATF	First latitude of data band (units 0.01°, $0 \leqslant$ LATF $\leqslant 18000$, South Pole = 0)
3	LATL	Last longitude of data band
4	LONF	First longitude of data band (units 0.01°, $0 \leqslant$ LONF $\leqslant 36000$, Greenwich = 0)
5	LONL	Last longitude of data band

The following 13 words contain the start locations (in bits) of the constituent vectors relative to the beginning of the record. The mnemonics and descriptions refer to these vectors.

Word	Mnemonic of vector	Description of vector
6	IDENT	80 bit character identifiers
7	LAT	Latitudes - 16 bit integers
8	LON	Longitudes - 16 bit integers
9	ITIM	Time displacement in minutes from assimilation time + 16383 ($0 \leqslant$ ITIM $\leqslant 32766$) - 16 bit integers
10	ITYP	Observation type codes - 8 bit integers
11	IQUAL	Quality indicators (range depending on ITYP) - 8 bit integers
12	M	Scaling factors for reduction of errors - 8 bit integers

Word	Mnemonic of vector	Description of vector
13	BITVARP	Indicators to signify presence of surface pressure (P_S) in observation - single bit indicators
14	BITVART	Indicators to signify presence of temperature (T) in observation - single bit indicators
15	BITVARU	Indicators to signify presence of velocity components (u, v) in observation - single bit indicators
16	BITVARR	Indicators to signify presence of humidity (r) in observation - single bit indicators
17	LOBDAT	No. of data elements in each observation - 8 bit integers
18	LOBPLV	No. of pressure levels in each observation - 8 bit integers

Words 19 et seq. contain the constituent vectors.

FLAG data records

As in the data records described above, there is a header section of 17 full word integers to provide the start locations in bits of the constituent vectors. These vectors are of single bits apart from the last as indicated. The data refer to all the pressure levels of each observation arranged consecutively.

Word	Mnemonic of vector	Description of vector
1	BITP	Indicators for P_S at each pressure level
2	BITT	Indicators for T at each pressure level
3	BITU	Indicators for u and v at each pressure level
4	BITR	Indicators for r at each pressure level
5	BITCHK1P	Flags set during 1st quality control check of P_S (or by data bank check)

Word	Mnemonic of vector	Description of vector
6	BITCHK1T	Flags set during 1st quality control check of T (or by data bank check)
7	BITCHK1U	Flags set during 1st quality control check of u and v (or by data bank check)
8	BITCHK1R	Flags set during 1st quality control check of r (or by data bank check)
9 to 12	BITCHK2P etc	Flags set during 2nd quality control check
13 to 16	BITFINP etc	Final quality control flags
17	IP	Pressure levels of observations (units 0.1mb) - 16 bit integers

Words 18 et seq. contain constituent vectors.

Notes: BITCHK1x flags are initially off except for data flagged by the data bank.

BITCHK2x flags are initially off.

BITFINx flags are initially on, unless intervention has insisted on acceptance of data.

DATA data records

The observations at each pressure level of each observation are arranged consecutively as 32 bit reals. The data at each level is ordered according to the heirarchy P_S,T,u,v,r. e.g.

$OB_{i,j,1}$	$OB_{i,j,2}$		$OB_{i,j+1,1}$	$OB_{i,j+1,2}$		$OB_{i+1,1,1}$	

$OB_{i,j,k}$ contains the k^{th} variable at the j^{th} level of the i^{th} observation. Different variables at each level are grouped according to the above heirarchy.

$OB_{i,1,1}$ is the first datum of the i^{th} observation and

$OB_{i,j,1}$ is the first datum at level j.

b. Observational data files DOER,DPER,INCR and FLAG2

DOER, DPER and INCR have precisely the same format as DATA.

FLAG2 has precisely the same format as FLAG

c. Model files BACK and YFE5DEG

The format of BACK is described in the companion note on the forecast model (FC).

YFE5DEG

i. Header record

At present YFE5DEG has no header record, though it is intended that it should be introduced with the same format described for the observational data sets. Adoption of the header would enable the following description to be generalised.

ii. Data records

The model errors are specified on a 5 degree latitude-longitude grid with 6 pressure levels in the vertical (5 for humidity). There are therefore 37 records for a global data set (row 1 being the North Pole), and 30 vectors of length 72 in each record defining the errors at 5 degree intervals. The first grid point of this coarse grid lies on the Greenwich meridian. The vectors are all contiguous and consist of 64 bit real numbers. The vectors are arranged in the following order:

$$P_S,\ r_1 \rightarrow r_5,\ T_1 \rightarrow T_6,\ u_1 \rightarrow u_6,\ v_1 \rightarrow v_6$$

where the levels 1 to 6 are 900 mb, 700 mb, 500 mb, 300 mb, 100 mb and 50 mb respectively.

d. Merged observational file OBDATA

i. Header record

This has the same format as that described for the observational data files.

ii. Data records

These records each contain 1 or 4 sub-records, and a header section is incorporated at the start of each record to facilitate access of these sub-records. The sub-secords all start on 64 bit full word boundaries and the header section contains their start locations (in words) relative to the beginning of the record.

Word	Description
1	No. of sub-records (1 if no data, 4 if data present)
2	Start location of 1st sub-record (INDEX record)
3	Start location of 2nd sub-record (FLAG record)
4	Start location of 3rd sub-record (DATA record)
5	Start location of 4th sub-record (BLWTS record)

Words 6 et seq. contain the sub-records.

The formats of the INDEX, FLAG and DATA records have already been described. The BLWTS record contains the weights and pointers necessary to enable model values to be interpolated to model grid points and is described below.

BLWTS sub-record

This sub-record contains 8 vectors and a header section is incorporated at the beginning to facilitate access of these vectors. The lengths of the vectors are the same as in the INDEX record. The vectors all start on 64 bit full word boundaries and the header section contains their displacement (in words) from the start of the sub-record. The mnemonics and descriptions refer to these vectors.

Word	Mnemonic of vector	Description of vector
1	IPRWS	Model row of pTr point immediately to north of observation - 16 bit integers
2	IURWS	Model row of uv point immediately to north of observation - 16 bit integers
3	IPPTS	Model pTr point number immediately to west of observation - 16 bit integers
4	IUPTS	Model uv point number immediately to west of observation - 16 bit integers
5	PAWTS	Distance in E-W direction to pTr point immediately to west of observation - 32 bit reals
6	UAWTS	Distance in E-W direction of uv point immediately to west of observation - 32 bit reals
7	PBWTS	Distance in N-S direction to pTr point immediately to north of observation - 32 bit reals
8	UBWTS	Distance in N-S direction to uv point immediately to north of observation - 32 bit real

Notes: The start addresses are in terms of displacements which are one less than the locations used hitherto.

pTr and uv points refer to the stagger of the variables on the model grid (see Annex 4).

"Distance in E-W direction" refers to the fraction of a grid-length in the East-West direction.

"Distance in N-S direction" refers to the fraction of a grid-length in the North-South direction.

e. Optimum interpolation weights file WEIGHTS

i. Header record

This record is 1 sector (512 words) long and contains the lengths of the subsequent data records in words 101 et seq. The number of records (excluding the header) is equal to the number of rows of the analysis grid.

It is intended that the more general header described previously should eventually be adopted.

ii. Data records

These records each contain 0 or 3 sub-records, and a header section is incorporated at the start of each record to facilitate access of these sub-records. The sub-records all start on 64 bit full word boundaries, and the header section contains their start locations (in words) relative to the beginning of the record.

Word	Description
1	No. of sub-records (Ø if no weights, 3 if weights present)
2	Start location of 1st sub-record (GATHER record)
3	Start location of 2nd sub-record (SCATTER record)
4	Start location of 3rd sub-record (WEIGHTS record)

GATHER sub-record

This sub-record contains 2 vectors and a header section containing the start location (in words) of these vectors relative to the beginning of the sub-record. Each vector starts on a 64 bit full word boundary.

Word	Description
1	Start location of 1st vector (IROW)
2	Start location of 2nd vector (IPOINT)

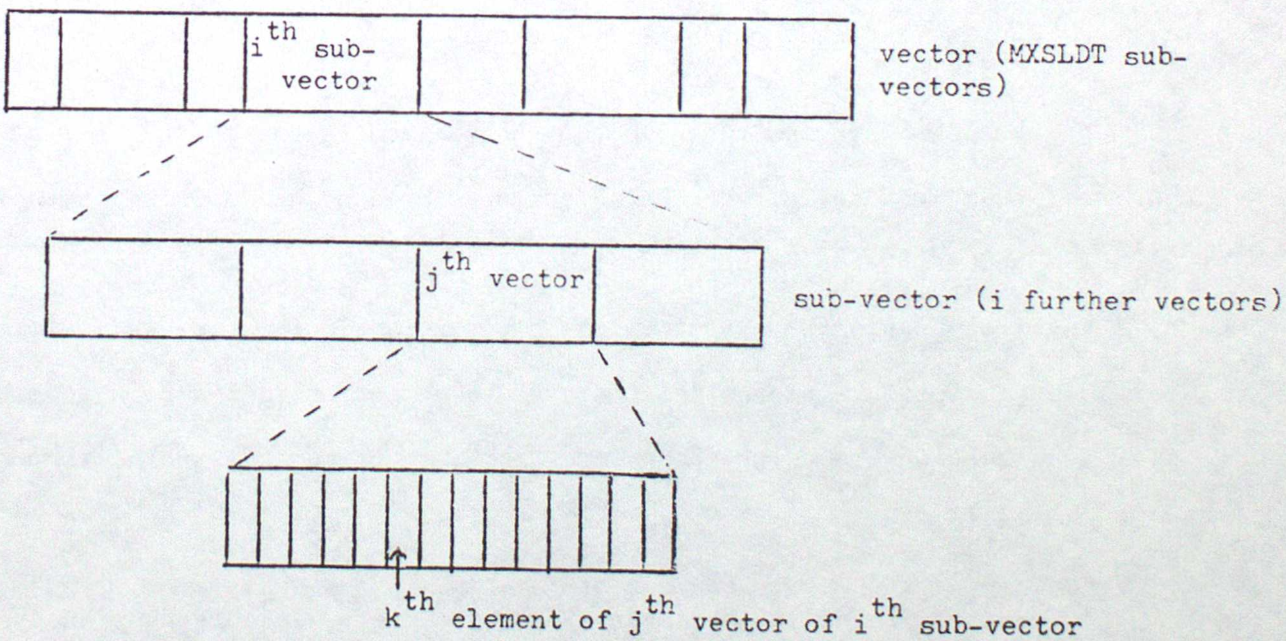
The remaining words of the header section contain information necessary to the deciphering of these vectors.

Word	Mnemonic	Description
3	MXSLDT	Max. number of observational data that can update a grid point.
3+ i ($1 \leq i \leq$ MXSLDT)	NWLOC(i)	Number of grid points updated by i observational data.

Words 4 + MXSLDT et seq. contain the vectors.

The organisation of the IROW and IPOINT vectors is the same and enables the location of an individual item of data to be defined. An element of the IROW vector contains the data row number containing the item as an 8 bit integer, whilst the corresponding element of IPOINT contains its location within that row (i.e. the location in the relevant record of the DATA data file) as a 16 bit integer.

The actual organisation of the vectors is rather complex to enable efficient vector processing. Each vector is divided into MXSLDT contiguous sub-vectors, the i^{th} sub-vector giving the data locations for those grid points which have i data updating them. Each sub-vector is in turn divided into i contiguous vectors, the j^{th} vector giving the locations of the j^{th} data updating the grid points. Therefore the k^{th} element of the j^{th} vector of the i^{th} sub-vector gives the location of the j^{th} datum updating the k^{th} grid point of that set of NWLOC(i) grid points which have i observational data updating them by optimal interpolation (see diagram below)



SCATTER sub-record

This sub-record contains 3 vectors and a header section containing the start location (in words) of these vectors relative to the beginning of the sub-record. Each vector starts on a 64 bit full word boundary.

Word	Description
1	Start location of 1st vector (ANLEV)
2	Start location of 2nd vector (PSTAR)
3	Start location of 3rd vector (ISCAT)

The remaining words of the header section contain information necessary to define the analysis grid points for the row.

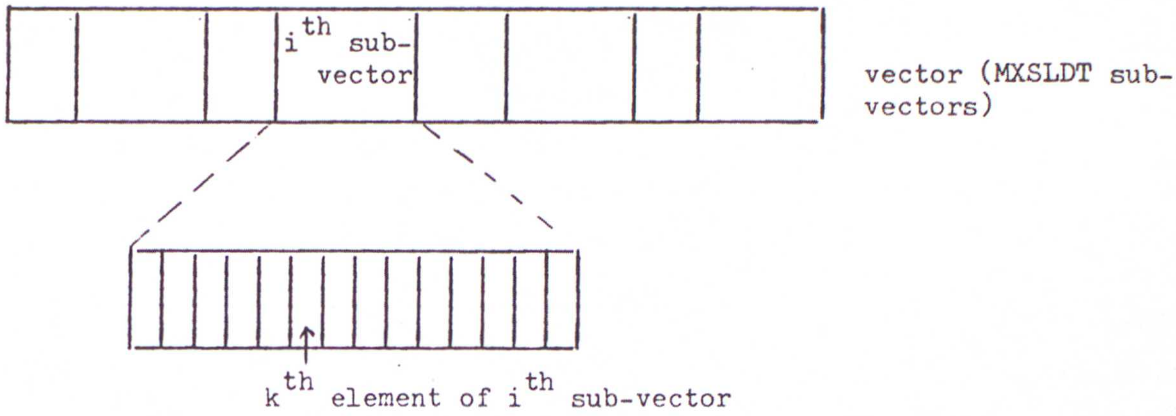
Word	Mnemonic	Description
4	RA3	Longitude of 1st point of analysis grid
5	NANLEV	Number of analysis levels
6	NANLON	Number of analysis grid points

Words 7 et seq. contain the vectors.

The ANLEV vector contains the ANLEV sigma levels of the analysis grid as 32 bit reals. The PSTAR vector contains the NALON surface pressure values of the background field as 32 bit reals. These two vectors then enable the pressure levels of the analysis grid, defined in the optimum interpolation stage, to be reconstructed (see section 3), these levels being those for which the optimum interpolation weights are valid.

The remaining ISCAT vector of 16 bit integers enables the analysis grid point locations of interpolated values to be identified. It consists of MXSLDT contiguous sub-vectors of which the i^{th} consists of NWLOC(i) elements. Thus the k^{th} element of the i^{th} sub-vector gives the analysis grid point location of the k^{th} grid point of that set of

NWLOC(i) grid points which have i observational data updating them (see diagram below and refer also to description of GATHER sub-record).



WEIGHTS sub-record

This sub-record has no header section and consists of just one vector. Each element of the vector is the optimal interpolation weight associated with the item of data located by the corresponding elements of the IROW and IPOINT vectors of the GATHER sub-record. The weights are stored as 16 bit integers having first been scaled by 10000. The sign of the weight is determined by whether the last bit is on or not. Thus only weights between 6.5535 and -6.5535 can be held in the file.

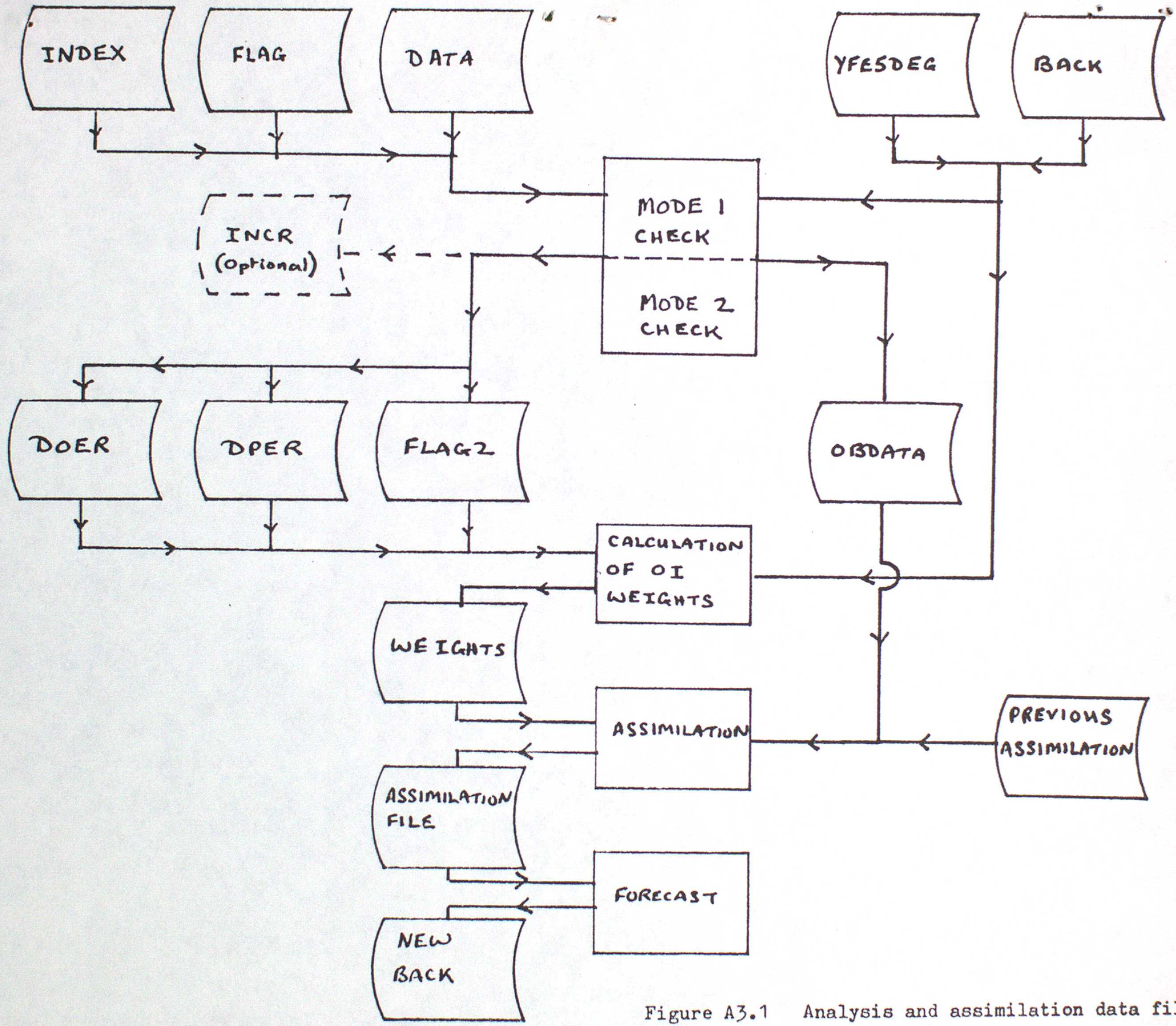


Figure A3.1 Analysis and assimilation data files.

Annex 4   Grids

There are four grids used in the data assimilation scheme and these are described below. The first concerns observational data and can be regarded as a quasi-random grid.

i. Data grid

Data are arranged in bands of $1\frac{1}{2}^{\circ}$ latitude width starting at the North Pole. Within the observational data files (see Annex 3), the observations are approximately ordered in terms of increasing longitude. This ordering is not exact. Within each observation the pressure levels are ordered from the bottom to top of the atmosphere. This ordering is exact except for, on occasions, the bottom levels of an upper ascent, when surface data may occur at a higher level than the lowest levels. This is due to extrapolation of data to model orography (see remarks in section 2).

ii. Model grid

Model data are defined on a latitude-longitude grid of grid length $\Delta\phi = 1.5^{\circ}$ in the North-South direction, and $\Delta\lambda = 1.875^{\circ}$ in the East-West direction. The variables are staggered on what is termed the Arakawa 'B' grid. The points at which the wind components are defined (uv points) are staggered half a grid length South and East of the points at which surface pressure, potential temperature and humidity are defined (pTr points) (see figure A4.1)

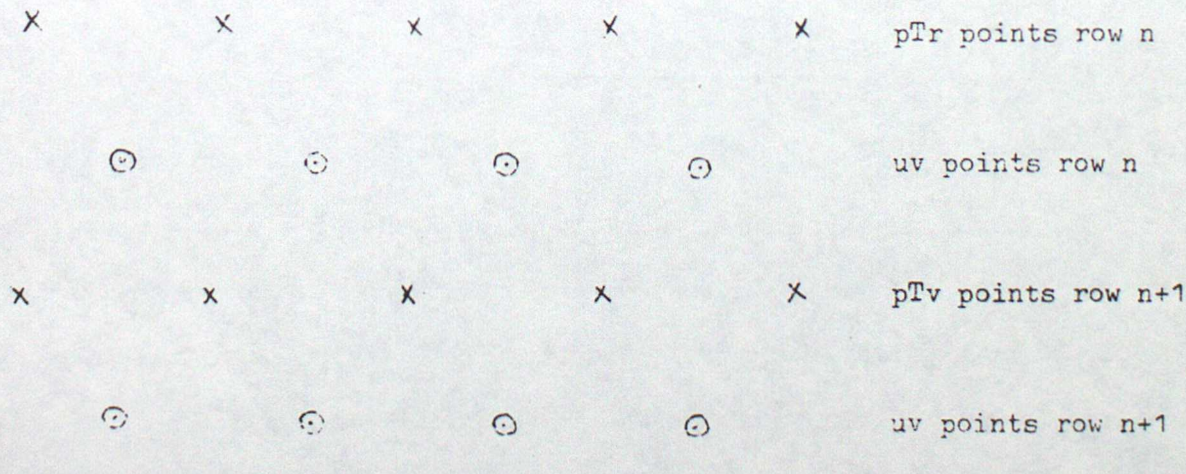


Figure A4.1   Stagger of model variables in horizontal

The first pTr row is at the North Pole and the first pTr point within each row is on the Greenwich meridian. The data are defined on 15 sigma levels in the vertical given in table A4.1.

Table A4.1 Model sigma levels

Level	Sigma
1	.997
2	.975
3	.935
4	.87
5	.79
6	.69
7	.59
8	.49
9	.39
10	.31
11	.25
12	.19
13	.125
14	.065
15	.025

iii. Model error grid

The model errors are defined on a latitude-longitude grid of 5° resolution in both the North-South and East-West direction with no stagger of the variables. The first row is at the North Pole and the first grid point within each row is on the Greenwich meridian. The data are defined on 6 pressure levels in the vertical given in table A4.2.

Table A4.2 Model error pressure levels

Level	Pressure (mb)
1	900
2	700
3	500
4	300
5	100
6	50

iv. Analysis grid

The analysis grid is of mixed type. Between 50N and 50S it is a latitude-longitude grid of the same resolution as the model grid. Poleward of these latitudes, where model data is filtered during an integration, the grid is quasi-homogeneous. The grid length remains 1.5° in the North-South direction, but in the East-West direction it retains approximately the same resolution as at 50°, subject only to the grid length being an integral fraction of 360°, and the number of grid points being at least 16. The numbers of grid points in each row and their latitudes are given in table A4.3. The variables are not staggered, but the first row is a half a grid length south of the North Pole and the first point within each row is on the Greenwich meridian. Because of this, interpolation of increments from the analysis grid to the model grid (see section 4.3) will remove two-grid length waves in the east-west direction for wind increments and will remove two-grid length waves in the north-south direction for increments of surface pressure, temperature and humidity. Strictly this only applies between 50N and 50S. In the vertical the analysis grid has the same 15 sigma levels as the model grid.

Table A4.3 Numbers of grid points in each analysis row and their latitudes

Row No.	|Latitude|	No. of points	Row No.	|Latitude|	No. of points
1 & 120	89.25	16	15 & 106	68.25	111
2 & 119	87.75	16	16 & 105	66.75	118
3 & 118	86.25	20	17 & 104	65.25	125
4 & 117	84.75	27	18 & 103	63.75	132
5 & 116	83.25	35	19 & 102	62.25	139
6 & 115	81.75	43	20 & 101	60.75	146
7 & 114	80.25	51	21 & 100	59.25	153
8 & 113	78.75	58	22 & 99	57.75	159
9 & 112	77.25	66	23 & 98	56.25	166
10 & 111	75.75	74	24 & 97	54.75	172
11 & 110	74.25	81	25 & 96	53.25	179
12 & 109	72.75	89	26 & 95	51.75	185
13 & 108	71.25	96	27 & 94	50.25	191
14 & 107	69.75	103	28 to 93	$\leqslant$ 48.75	192

Annex 5 **Finite Difference Approximations**

a. **Finite difference operators**

Let $\psi(\lambda, \phi, \sigma)$ be a field variable where λ, ϕ and σ are the longitude, latitude and vertical co-ordinate respectively. The finite difference operators are defined as follows:

$$\delta_\lambda \psi = \left[\psi(\lambda + \Delta\lambda/2) - \psi(\lambda - \Delta\lambda/2) \right] / \Delta\lambda$$

$$\delta_\phi \psi = \left[\psi(\phi + \Delta\phi/2) - \psi(\phi - \Delta\phi/2) \right] / \Delta\phi$$

$$\overline{\psi}^{\lambda} = \left[\psi(\lambda + \Delta\lambda/2) + \psi(\lambda - \Delta\lambda/2) \right] / 2$$

$$\overline{\psi}^{\phi} = \left[\psi(\phi + \Delta\phi/2) + \psi(\phi - \Delta\phi/2) \right] / 2$$

where $\Delta\lambda, \Delta\phi$ are the respective grid lengths.

b. **Geostrophic wind increments**

The finite difference form of equations (4.4) is

$$\Delta u_k = -\frac{1}{af} \left\{ \overline{\delta_\phi \Delta \Phi_k}^{\lambda} + c_p \sigma_k^{\kappa} \overline{\overline{\Delta\Theta_k}^{\lambda}}^{\phi} \overline{\delta_\phi \Pi}^{\lambda} + c_p \sigma_k^{\kappa} \overline{\overline{\Theta_k}^{\lambda}}^{\phi} \overline{\delta_\phi \Delta\Pi}^{\lambda} \right\}$$

$$\Delta v_k = -\frac{1}{af(\cos \phi)^{v}} \left\{ \overline{\delta_\lambda \Delta \Phi_k}^{\phi} + c_p \sigma_k^{\kappa} \overline{\overline{\Delta\Theta_k}^{\lambda}}^{\phi} \overline{\delta_\lambda \Pi}^{\phi} + c_p \sigma_k^{\kappa} \overline{\overline{\Theta_k}^{\lambda}}^{\phi} \overline{\delta_\lambda \Delta\Pi}^{\phi} \right\}$$

$$\Delta \Phi_k = R\,\Delta\Pi \left\{ \sum_{l=1}^{k-1} \sigma_l^{\kappa} \Theta_l \ln(\sigma_{l-\frac{1}{2}}/\sigma_{l+\frac{1}{2}}) + \sigma_k^{\kappa} \Theta_k \ln(\sigma_{k-\frac{1}{2}}/\sigma_k) \right\}$$
$$+ R\Pi \left\{ \sum_{l=1}^{k-1} \sigma_l^{\kappa} \Delta\Theta_l \ln(\sigma_{l-\frac{1}{2}}/\sigma_{l+\frac{1}{2}}) + \sigma_k^{\kappa} \Delta\Theta_k \ln(\sigma_{k-\frac{1}{2}}/\sigma_k) \right\}$$

where $(\cos \phi)^{v}$ refers to the cosine of latitude evaluate at relevant u,v point of the model grid (see Annex 4).

c. <u>Divergence damping</u>

The divergence D is defined as

$$D = \frac{1}{a\cos\phi}\left[\frac{\partial u}{\partial\lambda} + \frac{\partial}{\partial\phi}(v\cos\phi)\right]$$

and its finite difference approximation is

$$D = \frac{1}{a(\cos\phi)^{P}}\left[\overline{\delta_{\lambda} u}^{\phi} + \overline{\delta_{\phi} v(\cos\phi)^{v}}^{\lambda}\right]$$

where $(\cos\phi)^{P}$ refers to the cosine of latitude at a pTr point.

The damping terms in equations (4.7) at each level are then

$$\frac{\mu}{a\cos\phi}\frac{\partial D}{\partial\lambda} = \frac{\mu}{a(\cos\phi)^{v}}\overline{\delta_{\lambda} D}^{\phi}$$

$$\frac{\mu}{a}\frac{\partial D}{\partial\phi} = \frac{\mu}{a}\overline{\delta_{\phi} D}^{\lambda}$$

Annex 6 Observational error correlations

At present only temperature data derived from satellite observations (SATEMs) are assumed to have correlated errors. The correlation function for these is defined as

$$\mu^{o}_{ij} = e^{-p\theta^{2}_{ij} - q\,\left|\ln(p_i/p_j)\right|}$$

where θ_{ij} is the angle in radians subtended at the Earth's centre by points i and j and p_i and p_j are their pressure levels. The constants p and q are defined in Annex 10. No correlation of errors in the vertical for data from a single upper air observation is assumed at present. This is only required if the maximum number of pressure levels (M_3) from which data may be selected from a single observation exceeds unity (see section 3.5 and Annex 10).

Annex 7 Background field error correlations

Let $\mathcal{E}_i^x$ be the background error for the model variable x, E_i^x its root mean square value, and $\mu(x_i, y_i)$ the correlation between the errors of variables x and y at points i and j respectively. Then equation (3.9) for the error covariances becomes

$$\langle \mathcal{E}_i^x \mathcal{E}_j^y \rangle = E_i^x \mu(x_i, y_j) E_j^y$$

The functional forms of the error covariances between the variables of surface pressure(p_s), potential temperature (θ) and the westerly (u) and southerly (v) wind components are all derived from the single height-height error correlation function defined in equation (3.10). This is achieved by enforcing hydrostatic and geostrophic constraints on the error fields. Thus if $\emptyset$ is the geopotential height

$$\left.\begin{aligned}
\frac{\mathcal{E}_i^p}{p_i} &= -\frac{\mathcal{E}_i^{\emptyset}}{RT_i} \quad \text{(isothermal atmosphere)} \\
\mathcal{E}_i^{\theta} &= \frac{1000^K}{R} e^{-K\ln p_i} \frac{\partial \mathcal{E}_i^{\emptyset}}{\partial \ln p_i} \\
\mathcal{E}_i^{u_x} &= \frac{G_i}{f_i} \frac{\partial \mathcal{E}_i^{\emptyset}}{\partial y} \\
\mathcal{E}_i^{u_y} &= \frac{-G_i}{f_i} \frac{\partial \mathcal{E}_i^{\emptyset}}{\partial x}
\end{aligned}\right) \qquad \text{(A7.1)}$$

where x and y are local Cartesian co-ordinates, u_x and u_y the respective velocity components, f the Coriolis parameter and G defines the degree of geostrophic balance (ie G = 1 corresponds to geostrophic balance). To enable the results to apply in the Equatorial region it is assumed that the ratio G/f remain finite and non-zero in that area.

The following formulae for calculating univariate covariance functions can then be derived:

$$\langle \mathcal{E}_i^p \, \mathcal{E}_j^p \rangle = \frac{p_i p_j}{R^2 T_i T_j} \langle \mathcal{E}_i^\emptyset \, \mathcal{E}_j^\emptyset \rangle \text{(isothermal atmosphere)}$$

$$\langle \mathcal{E}_i^\theta \, \mathcal{E}_j^\theta \rangle = \frac{1000^{2K} e^{-K\ln(p_i p_j)}}{R^2} \quad \frac{\partial^2 \langle \mathcal{E}_i^\phi \, \mathcal{E}_j^\phi \rangle}{\partial \ln p_i \, \partial \ln p_j}$$

$$\langle \mathcal{E}_i^{u_x} \, \mathcal{E}_i^{u_x} \rangle = \frac{G_i G_j}{f_i f_j} \, \frac{\partial^2 \langle \mathcal{E}_i^\emptyset \, \mathcal{E}_j^\emptyset \rangle}{\partial y_i \partial y_j} \qquad \text{(A7.2)}$$

$$\langle \mathcal{E}_i^{u_x} \, \mathcal{E}_j^{u_y} \rangle = \frac{-G_i G_j}{f_i f_j} \, \frac{\partial^2 \langle \mathcal{E}_i^\emptyset \, \mathcal{E}_j^\emptyset \rangle}{\partial y_i \, \partial x_j}$$

$$\mathcal{E}_i^{u_y} \, \mathcal{E}_j^{u_x} = \frac{-G_i G_j}{f_i f_j} \, \frac{\partial^2 \langle \mathcal{E}_i^\emptyset \, \mathcal{E}_j^\emptyset \rangle}{\partial x_i \, \partial y_j}$$

$$\mathcal{E}_i^{u_y} \, \mathcal{E}_j^{u_y} = \frac{G_i G_j}{f_i f_j} \, \frac{\partial^2 \langle \mathcal{E}_i^\emptyset \, \mathcal{E}_j^\emptyset \rangle}{\partial x_i \, \partial x_j}$$

If the pressure-pressure error correlations are confined to surface pressure then the assumption of an isothermal atmosphere is well justified. It is now further assumed that:

$$E_i^{p_s} = E_i^\emptyset \, \frac{P_{si}}{RT_{si}}$$

$$E_i^\theta = E_i^\emptyset \, \frac{1000^K \, e^{-K\ln p_i}}{R} \sqrt{2b} \qquad \text{(A7.3)}$$

$$E_i^u = E_i^{u_x} = E_i^{u_y} = E_i^\emptyset \, \frac{G_i}{f_i} \sqrt{2a}$$

where the parameters a and b are those used in equation (3.10) defining the height-height error correlation function. Note that specifying E^u and E^{ϕ} from separate statistics avoids specification of G and making assumptions regarding the degree of geostrophic balance. The error covariances can now be derived as

$$\left.\begin{aligned}
\langle \varepsilon_i^{\phi}\, \varepsilon_j^{\phi} \rangle &= E_i^{\phi}\, E_j^{\phi}\, \mu(\phi_i, \phi_j) \\
\langle \varepsilon_i^{p_s}\, \varepsilon_j^{p_s} \rangle &= E_i^{p_s}\, E_j^{p_s}\, \mu(\phi_i, \phi_j) \\
\langle \varepsilon_i^{\theta}\, \varepsilon_j^{\theta} \rangle &= E_i^{\theta}\, E_j^{\theta} \left[1-2b \ln^2(p_i/p_j) \right] \mu(\phi_i, \phi_j) \\
\langle \varepsilon_i^{u_x}\, \varepsilon_j^{u_x} \rangle &= E_i^{u}\, E_j^{u} \left[1-2a\, (y_i - y_j)^2 \right] \mu(\phi_i, \phi_j) \\
\langle \varepsilon_i^{u_x}\, \varepsilon_j^{u_y} \rangle = \langle \varepsilon_i^{u_y}\, \varepsilon_j^{u_x} \rangle &= 2a\, E_i^{u} E_j^{u} (x_i - x_j)(y_i - y_j)\, \mu(\phi_i, \phi_j) \\
\langle \varepsilon_i^{u_y}\, \varepsilon_j^{u_y} \rangle &= E_i^{u} E_j^{u} \left[1-2a (x_i - x_j)^2 \right] \mu(\phi_i, \phi_j)
\end{aligned}\right) \qquad \text{(A7.4a)}$$

where $\mu(\phi_i, \phi_j)$ is defined by equation (3.10) and $r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2$. To derive the covariances for the westerly (u) and southerly (v) components of wind refer to figure A7.1 where the local x axis is drawn along the line joining points i and j. ie. $y_i = y_j$ and $r_{ij} = x_j - x_i$.

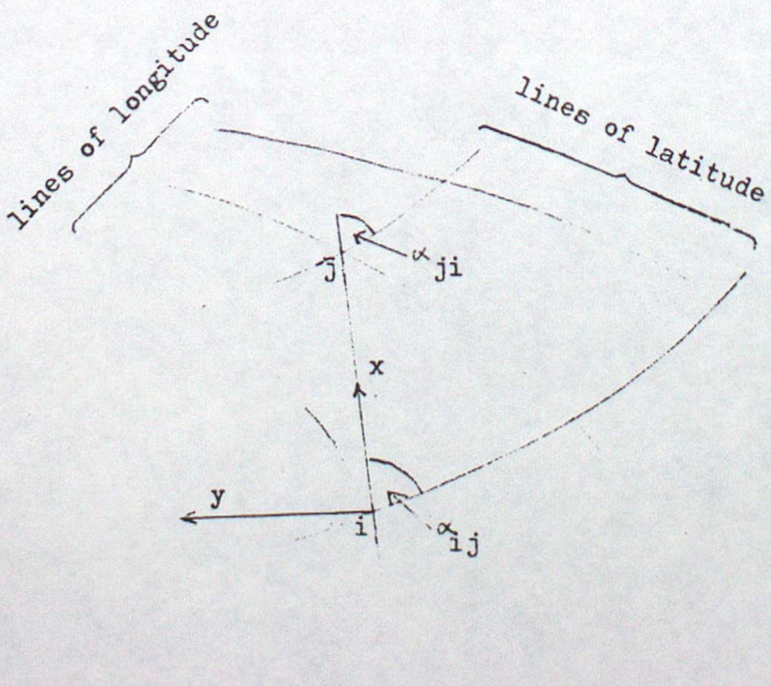


Figure A7.1

Then the covariances are related by

$$\langle \varepsilon_i^u \varepsilon_j^u \rangle = \cos\alpha_{ij} \cos\alpha_{ji} \langle \varepsilon_i^{u_x} \varepsilon_j^{u_x} \rangle + \sin\alpha_{ij} \sin\alpha_{ji} \langle \varepsilon_i^{u_y} \varepsilon_j^{u_y} \rangle$$

$$\langle \varepsilon_i^u \varepsilon_j^v \rangle = \cos\alpha_{ij} \sin\alpha_{ji} \langle \varepsilon_i^{u_x} \varepsilon_j^{u_x} \rangle - \sin\alpha_{ij} \cos\alpha_{ji} \langle \varepsilon_i^{u_y} \varepsilon_j^{u_y} \rangle$$

$$\langle \varepsilon_i^v \varepsilon_j^u \rangle = \sin\alpha_{ij} \cos\alpha_{ji} \langle \varepsilon_i^{u_x} \varepsilon_j^{u_x} \rangle - \cos\alpha_{ij} \sin\alpha_{ji} \langle \varepsilon_i^{u_y} \varepsilon_j^{u_y} \rangle$$

$$\langle \varepsilon_i^v \varepsilon_j^v \rangle = \sin\alpha_{ij} \sin\alpha_{ji} \langle \varepsilon_i^{u_x} \varepsilon_j^{u_x} \rangle + \cos\alpha_{ij} \cos\alpha_{ji} \langle \varepsilon_i^{u_y} \varepsilon_j^{u_y} \rangle$$

where α_{ij} is the angle between the vector $\underline{r}_{ij}$ and the latitude at point i. Thus the final velocity error covariances are

$$\left.\begin{aligned}
\langle \varepsilon_i^u \varepsilon_j^u \rangle &= E_i^u E_j^u \cos\alpha_{ij} \cos\alpha_{ji} \left[1 + \tan\alpha_{ij} \tan\alpha_{ji} (1-2ar_{ij}^2) \right] \mu(\phi_i, \phi_j) \\
\langle \varepsilon_i^u \varepsilon_j^u \rangle &= E_i^u E_j^u \cos\alpha_{ij} \cos\alpha_{ji} \left[\tan\alpha_{ji} - \tan\alpha_{ij} (1-2ar_{ij}^2) \right] \mu(\phi_i, \phi_j) \\
\langle \varepsilon_i^v \varepsilon_j^u \rangle &= E_i^u E_j^u \cos\alpha_{ij} \cos\alpha_{ji} \left[\tan\alpha_{ij} - \tan\alpha_{ji} (1-2ar_{ij}^2) \right] \mu(\phi_i, \phi_j) \\
\langle \varepsilon_i^v \varepsilon_j^v \rangle &= E_i^u E_j^u \cos\alpha_{ij} \cos\alpha_{ji} \left[\tan\alpha_{ij} \tan\alpha_{ji} + (1-2ar_{ij}^2) \right] \mu(\phi_i, \phi_j)
\end{aligned}\right\} \quad \text{(A7.4b)}$$

If the assumptions (A7.3) are not satisfied then the covariances defined in equations (A7.4a and b) will not be hydrostatically or geostrophically linked.

Although the optimum interpolation is univariate (indeed only u-u and v-v error correlationsare used equatorwards of 63 degrees), it is clearly possible to derive multi-variate covariances in a similar fashion. For completeness these are detailed below:

$$
\left.\begin{aligned}
\langle \varepsilon_i^{p_s} \varepsilon_j^{\theta} \rangle &= E_i^{p_s} E_j^{\theta} \sqrt{2b}\, \ln (p_i/p_j)\, \mu(\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{\theta} \varepsilon_j^{p_s} \rangle &= - E_i^{\theta} E_j^{p_s} \sqrt{2b}\, \ln (p_i/p_j)\, \mu(\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{p_s} \varepsilon_j^{u} \rangle &= E_i^{p_s} E_j^{u} \sqrt{2a}\; r_{ij}\; \sin \alpha_{ji}\; (\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{u} \varepsilon_j^{p_s} \rangle &= - E_i^{u} E_j^{p_s} \sqrt{2a}\; r_{ij}\; \sin \alpha_{ij}\, \mu(\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{p_s} \varepsilon_j^{v} \rangle &= - E_i^{p_s} E_j^{u} \sqrt{2a}\; r_{ij}\; \cos \alpha_{ji}\, \mu(\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{v} \varepsilon_j^{p_s} \rangle &= E_i^{u} E_j^{p_s} \sqrt{2a}\; r_{ij}\; \cos \alpha_{ij}\, \mu(\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{\theta} \varepsilon_j^{u} \rangle &= - E_i^{\theta} E_j^{u}\; 2\sqrt{ab}\; r_{ij}\; \ln(p_i/p_j)\; \sin \alpha_{ji}\, \mu(\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{u} \varepsilon_j^{\theta} \rangle &= - E_i^{u} E_j^{\theta}\; 2\sqrt{ab}\; r_{ij}\; \ln(p_i/p_j)\; \sin \alpha_{ij}\, \mu(\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{\theta} \varepsilon_j^{v} \rangle &= E_i^{\theta} E_j^{u}\; 2\sqrt{ab}\; r_{ij}\; \ln(p_i/p_j)\; \cos \alpha_{ji}\, \mu(\emptyset_i, \emptyset_j) \\
\langle \varepsilon_i^{v} \varepsilon_j^{\theta} \rangle &= E_i^{u} E_j^{\theta}\; 2\sqrt{ab}\; r_{ij}\; \ln(p_i/p_j)\; \cos \alpha_{ij}\, \mu(\emptyset_i, \emptyset_j)
\end{aligned}\right\} \quad \text{(A7.5)}
$$

Annex 8 Observational error levels

Root mean square observational errors are tabulated within the optimal interpolation program according to observation type and pressure level. Values are given in Table A8.2 where, since certain observations share the same error levels, the observations have been assigned to groups. The composition of these groups is given in Table A8.1, together with the type number of the observation as given in the vector ITYP of the INDEX data file (see Annex 3). Surface data (ie data from a level which has a report of surface pressure) are given their own set of error levels in the table. Data from an observation whose type the program does not recognise is assigned an error level of 99. Data with this expected error level are ignored by the program.

Table A8.1 Data types and groups

Type No.	Description	Group
1	BOGUS	A
2	TEMP	B
3	TEMPSHIP	B
4	TEMPDROP (Drop-sonde)	B
5	ROCOB (Rocket-sonde)	B
6	ROCOB SHIP	B
7	PILOT	B
8	PILOTSHIP	B
9	SATEM	D
10	SYNOP	B
11	SYNOPAUTO	B
12	SHIP (surface)	C
13	SHIPAUTO (surface)	C
14	OWS (surface)	B
15	AIREP	E
16	COLBA (constant level balloon)	I
17	DRIBU (Drifting buoy)	G
18	SATOB	F
19	PAOB	H

Table A8.2 Observational errors

Surface pressure (mb)

Group								
A	B	C	D	E	F	G	H	I
1.0	1.7	2.0	99	99	99	4.0	2.5	99

Potential temperature (K)

Pressure level (mb)	Group A	B	C	D	E	F	G	H	I
Surface	99	99	99	99	99	99	99	99	99
1100	0.8	1.2	99	2.7	99	99	99	1.8	99
1000	0.8	1.2	99	2.7	99	99	99	1.8	99
850	0.9	1.3	99	2.5	99	99	99	2.0	99
700	1.0	1.4	99	2.5	3.0	99	99	2.1	99
500	1.2	1.5	99	2.5	3.0	99	99	2.3	99
400	1.3	1.6	99	2.0	3.5	99	99	2.4	99
300	1.4	1.7	99	1.7	3.5	99	99	2.6	99
250	1.5	1.8	99	1.8	4.0	99	99	2.7	3.0
200	1.6	1.9	99	1.9	4.3	99	99	2.9	3.5
150	1.8	2.1	99	2.1	4.7	99	99	3.2	4.0
100	2.3	2.7	99	2.7	5.0	99	99	4.1	4.5
70	3.2	3.6	99	3.6	6.0	99	99	5.4	5.0
50	3.7	4.0	99	5.0	99	99	99	6.0	99
30	4.3	4.6	99	5.1	99	99	99	6.9	99
20	5.0	5.2	99	5.2	99	99	99	7.8	99
10	6.3	6.3	99	6.3	99	99	99	9.5	99

nd component (ms^{-1})

Pressure level (mb)	A	B	C	D	E	F	G	H	I
Surface	2.0	99	4.5	99	99	99	99	99	99
1100	0.6	1.0	99	99	99	5.1	99	99	99
1000	0.6	1.0	99	99	99	5.1	99	99	99
850	0.6	1.0	99	99	99	4.8	99	99	99
700	0.8	1.0	99	99	7.0	4.0	99	99	99
500	1.0	1.5	99	99	7.0	4.0	99	99	99
400	1.3	2.0	99	99	7.0	8.0	99	99	99
300	1.5	2.5	99	99	7.0	10.0	99	99	99
250	1.7	2.5	99	99	7.0	11.0	99	99	7.0
200	2.0	3.0	99	99	7.0	11.0	99	99	7.0
150	2.0	3.0	99	99	7.0	10.0	99	99	7.0
100	2.0	3.0	99	99	7.0	10.0	99	99	7.0
70	2.0	3.0	99	99	7.0	10.0	99	99	7.0
50	2.0	3.0	99	99	7.0	10.0	99	99	99
30	2.0	3.0	99	99	7.0	10.0	99	99	99
20	2.0	3.0	99	99	7.0	10.0	99	99	99
10	2.0	3.0	99	99	7.0	10.0	99	99	99

Humidity (gm/gm)

midity (gm/gm)

Pressure level (mb)	Group A	Group B	Group Other
Surface	99	99	99
1100	.0012	.0012	99
1000	.0012	.0012	99
850	.001	.001	99
700	.001	.001	99
500	.0005	.0005	99
400	.0003	.0003	99
300	.0001	.0001	99
250	.0001	.0001	99
200	.0001	.0001	99
150	.0001	.0001	99
100	.0001	.0001	99
70	.0001	.0001	99
50	.0001	.0001	99
30	.0001	.0001	99
20	.0001	.0001	99
10	.0001	.0001	99

Note the error quoted for wind is the component error i.e. vector wind error = $\sqrt{2}$ x component wind error.

Annex 9 Background field error levels

Root mean square background field errors (or the model errors of a six hour forecast) are tabulated on a coarse 5 degree latitude-longitude grid with 6 pressure levels in the vertical (see Annex 4 for details). Clearly all these values cannot be reproduced here, and instead the maximum, minimum and mean values for each of the variables on selected rows and pressure levels are given in tables A9.1 to A9.4.

Table A9.1 Surface pressure errors (mb)

Lat	Error (max/min/mean)
90N	7.84/5.63/6.71
70N	6.02/2.91/4.29
50N	5.55/2.65/3.83
30N	5.26/2.52/3.83
10N	4.97/2.38/3.90
10S	5.21/2.78/4.42
30S	5.55/2.72/4.42
50S	6.34/4.01/5.36
70S	6.79/4.88/5.81
90S	8.95/6.43/7.66

Table A9.2 Potential temperature errors (K)

Lat	Errors (max/min/mean)		
	900mb	500mb	300mb
90N	3.07/2.20/2.63	6.69/4.80/5.72	9.07/6.51/7.76
70N	2.97/2.03/2.49	6.24/3.01/4.44	8.61/4.16/6.13
50N	2.52/1.71/2.09	4.56/2.18/3.15	8.04/3.85/5.56
30N	1.51/1.02/1.26	2.48/1.19/1.81	4.50/2.15/3.28
10N	1.21/0.85/1.03	2.16/1.09/1.72	4.03/2.04/3.21
10S	1.21/0.87/1.05	2.36/1.20/1.91	4.18/2.12/3.38
30S	1.55/1.07/1.31	3.07/1.51/2.45	5.00/2.45/3.99
50S	2.95/2.12/2.52	5.46/3.46/4.63	7.67/4.85/6.50
70S	3.74/2.69/3.20	6.24/4.48/5.34	8.44/6.06/7.22
90S	3.63/2.60/3.10	6.08/4.36/5.20	7.42/5.33/6.35

Table A9.3 Wind component errors (ms^{-1})

Lat	Errors (max/min/mean)		
	900mb	500mb	300mb
90N	6.22/4.46/5.32	11.46/8.22/9.80	13.42/9.64/11.48
70N	6.40/4.37/5.36	11.36/5.49/8.10	13.55/6.54/9.66
50N	6.66/4.51/5.51	10.20/4.88/7.05	15.53/7.43/10.73
30N	6.10/4.13/5.09	8.49/4.07/6.19	13.32/6.37/9.70
10N	5.79/4.07/4.96	8.73/4.43/6.97	14.11/7.16/11.24
10S	5.79/4.15/5.05	9.57/4.85/7.74	14.63/7.42/11.83
30S	6.28/4.35/5.31	10.53/5.16/8.39	14.80/7.25/11.79
50S	7.79/5.60/6.65	12.21/7.72/10.34	14.82/9.37/12.54
70S	8.07/5.79/6.90	11.38/8.17/9.73	13.29/9.54/11.37
90S	7.34/5.27/6.28	10.40/7.47/8.90	10.98/7.88/9.39

Table A9.4 Humidity errors (gm/gm x 10^4)

Lat	Errors (max/min/mean)		
	900mb	500mb	300mb
90N	7.24/5.20/6.20	1.81/1.30/1.55	0.60/0.60/0.60
70N	7.93/5.70/6.80	2.38/1.71/2.04	0.60/0.60/0.60
50N	14.63/10.45/12.48	2.98/2.13/2.55	0.60/0.60/0.60
30N	27.51/19.66/23.48	8.47/6.05/7.23	0.71/0.60/0.62
10N	31.51/22.51/26.92	12.38/8.85/10.58	1.41/1.01/1.21
10S	49.14/35.46/42.20	19.90/14.36/17.09	1.57/1.13/1.35
30S	36.45/26.17/31.15	10.49/7.53/8.96	1.22/0.88/1.04
50S	27.04/19.42/23.13	8.09/5.80/6.92	0.71/0.60/0.62
70S	16.18/11.62/13.84	4.77/3.43/4.08	0.71/0.60/0.62
90S	11.73/8.43/10.04	3.49/2.51/2.99	0.71/0.60/0.62

Note that the wind error quoted is the component error i.e. vector wind error = $\sqrt{2}$ x component wind error

Current values of analysis and assimilation parameters

Table A10.1 Data selection parameters (see section 3.4)

Parameter	Value	Description
Ri	$5\frac{1}{4}^{\circ}$ of latitude	Radius of influence
R_p	1.5	Pressure ratio defining vertical search extent
M_1	20	Maximum number of observations selected for each grid point position
M_2	7	Maximum number of data selected for each grid point variable
M_3	1	Maximum number of pressure levels of each complete observation from which data may be selected for each grid point variable
F(i)		Adjustment factors for variables
	0.0 (i = 1 to 9)	i = variable type, see table A8.1 for details
	2.0 (i = 10 to 14)	
	3.0 (i = 15, 16)	
	2.5 (i = 17,18)	
	1.0 (i = 19)	

Table A10.2 Quality control parameters (see sections 3.4 and 3.8)

Parameter	Value	Description
$n_1(i)$	12 (i = 1) 2 (i = 2 to 8,10,14 to 16) 2 (i = 9,11,12,13,15,17,18 & 19)	Number of standard deviations of expected error in Mode 1 criterion
$n_2(i)$	4 (i = 1) 3 (i = 2 to 8,10,14 & 16) 6 (i = 9,11,12,13,15,17,18 & 19)	Number of standard deviations of expected error in Mode 2 criterion

Table A10.3 Observational error correlation parameters (see Annex 6)

$p = 400$

$q = 1.83$

Table A10.4 Background error correlation parameters (see section 3.6)

$a = 3.3 \times 10^{-12}\ m^{-2}$ [equivalent to $(550.406\ km)^{-2}$]

$b = 3$

$c = 3.3 \times 10^{-12}\ m^{-2}$

$d = 3$

Table A10.5 Relaxation coefficients (see equation 4.1)

All values vary linearly in time from zero to final value given in table

Coefficient	Final Value	Description
	0.125	relaxation coefficient for all variables
	0.3	hydrostatic temperature increments coefficient
	0.0 between 30N and 30S	geostrophic wind increments coefficient
	0.075 levels 1 to 7	
	0.060 level 8	
	0.045 level 9	
	0.030 level 10	
	0.015 level 11	
	0.0 levels 12 to 15	

Table A10.6 Damping and diffusion coefficients (see section 4.1 and 4.7)

Coefficient	Value	Description
K_T	$2 \times 10^{14}\ m^4\ s^{-1}\ K^{-1}$	Non-linear diffusion coefficient for temperature
K_u	$2 \times 10^{14}\ m^3$	Non-linear diffusion coefficient for wind
K_r	$2 \times 10^{14}\ m^4\ s^{-1}$	Non-linear diffusion coefficient for humidity
μ	$5 \times 10^6\ m^2\ s^{-1}$	Divergence damping coefficient during assimilation
	$5 \times 10^6\ m^2\ s^{-1}$ decreasing to zero	Divergence damping coefficient during six hour forecast to produce background fields