CHAPTER 3

BACKGROUND TO COMPUTER MODELS

CHAPTER 3

BACKGROUND TO COMPUTER MODELS

Contents

LIST OF DIAGRAMS

LIST OF SYMBOLS

a	horizontal grid length on the polar stereographic map projection
c	speed of gravity waves
c_p	specific heat of air at constant pressure
C_D	drag coefficient
C_E	exchange coefficient for sensible and latent heat
D	divergence
f	Coriolis parameter
F_G	flux of heat into the ground
F_H, F_W	flux of heat, water vapour into atmosphere at 1000 mb
g	acceleration due to gravity
h	height of an isobaric surface
h'	thickness of 100-mb layer
H	topography height
$\mathbf{i},\mathbf{j},\mathbf{k}$	unit vectors in x, y, z directions
K	lateral diffusion coefficient
L	latent heat of condensation
m	map magnification factor
p	pressure
Q	rate of diabatic heating
r	humidity mixing ratio
r_s	saturation humidity mixing ratio
R	gas constant
R_N	net radiation flux at ground
T	thermodynamic temperature
u,v	horizontal velocity components
u^*, v^*	u/m, v/m
$\mathbf{V}$	horizontal wind velocity vector
w	vertical velocity, dz/dt
W	rate of change of humidity mixing ratio by evaporation, etc.
x, y	horizontal co-ordinates on polar stereographic map projection
z	height
Z	a meteorological variable
β	static stability parameter
Γ_d	dry-adiabatic lapse rate
δt	time-step of integration
Δp	vertical grid interval, 100 mb
θ	potential temperature
κ	R/c_p
μ	m^2
ρ	air density
$\boldsymbol{\tau}$	frictional stress
$\boldsymbol{\tau}^*$	$\boldsymbol{\tau}/m$
ϕ	latitude
χ	F_G/R_N
ω	vertical component of motion, dp/dt
∇	$\mathbf{i}(\partial/\partial x) + \mathbf{j}(\partial/\partial y)$
ζ	vorticity

Suffixes

k	pressure level index
o	indicates value at ground level
a	indicates value at 10 metres
10	indicates value at 1000 mb
9.5	indicates value at 950 mb
n	an integer

CHAPTER 3

BACKGROUND TO COMPUTER MODELS

3.1 INTRODUCTION

3.1.1 The requirement for numerical models

In setting out to produce a weather forecast the meteorologist is first of all faced with the task of assimilating a large quantity of data and reducing it to a form which is related in some way to the physical and dynamical processes occurring in the atmosphere - usually in terms of pressure patterns and fronts at the surface and of the contours of isobaric surfaces aloft. This is the analysis stage. He then has to estimate how the state of the atmosphere is likely to change during the forecast period, using his knowledge of the typical life histories of various types of feature, together with guide-lines and rules derived from simplified versions of the equations of motion. Finally, he must interpret his forecast charts to satisfy his customers' needs, sometimes by further processing, but often by the use of experience, physical reasoning or statistical relationships to predict the sequence of weather or the behaviour of individual elements at a given place.

It is hardly surprising that, when electronic computers appeared on the scene, attention was turned to possible ways of applying them to some of the above tasks. Developments in computer technology have been very rapid indeed, and so have those of the meteorological models, with new models often being derived in anticipation of advances in computer performance.

The virtues of the computer lie in its ability to process large amounts of data quickly and accurately according to precise instructions, and to carry out complex and laborious arithmetical manipulations at a great speed. Yet the quality of the products is determined entirely by the man-made procedures for handling the data and performing the calculations. In the analysis stage of a meteorological model, for example, the quality-control routines may reject a perfectly good observation or accept one which contains an error. This is because the computer model, however sophisticated, responds inflexibly to the input, whereas a human being, faced with the same observations, is able to allow himself greater freedom in exercising his judgement. Again, the forecast model may not reflect adequately all the complex physical and dynamical processes occurring in the atmosphere, or the approximations that have to be used in the treatment of the differential equations may lead to errors. The human forecaster can quickly learn the model's shortcomings after some experience of its use, and can apply some degree of correction to the computer products. For these reasons, most if not all meteorological services use a combination of human skills and computer techniques - the 'man-machine mix'. The relative amounts of man and machine in the aggregate vary, depending to a large extent on the characteristics of the model, but also on the attitude of the management of a particular organization and, to some degree, on the individual forecaster. The aim in all instances is to make the fullest and most efficient use of both the computer, with its capability for performing vast amounts of arithmetic, and the human being with his capacity for exercising judgement of a kind not yet programmable for the computer.

3.1.2 A brief history

The first attempt to produce a numerical forecast was made, long before electronic computers were dreamed of, by L. F. Richardson[1] in the early 1920s. In many ways his approach was similar to today's: in particular, he used the basic equations of motion for his calculations, and he introduced the effects of a number of physical processes such as radiation, condensation, etc. The forecast was a failure, partly because of the inadequacy of the data, but also because there were fundamental problems in the formulation of the model and in the mathematical procedures. However, it does appear that if the calculations could have been continued they might have settled down, after initial large oscillations, to more reasonable values. The production of forecasts in this way was far from a practicable proposition even if the products had been acceptable - it would have required a battery of some thousands of assistants merely to keep pace with actual events.

Richardson based his approach on the so-called 'primitive equations' which, as the name implies, are the basic ones such as the momentum equation relating the acceleration of the air to the pressure gradients. The disadvantage is that the equations describe atmospheric motions on all scales, including short-period gravity and sound waves, which are too small in scale to be dealt with directly by the numerical model, but which are likely to amplify rapidly and unrealistically and ruin the calculations unless very short time-steps are used in the integrations. A second disadvantage is that the accelerations (the terms needed to predict the development and movement of systems) are given as the small differences between larger terms.

These two major difficulties led to the primitive-equation approach being dropped for the first numerical models to be used operationally. Instead, use was made of the vorticity equation (section 2.2.5.1 of Chapter 2), relating changes in the component of vorticity about a vertical axis to the horizontal divergence of the motion. The relationship between the wind and the pressure field is required in this type of model; in early experiments the assumption was generally made that the flow was geostrophic, but more sophisticated approximations were made later. This technique, properly applied, ensured that the small-scale motions were filtered out (the models are often referred to as 'filtered models'), leaving only those features meteorologically significant on a scale of 1000 kilometres or more.

The main advantage of the filtered models was that the meteorologically important developments could be studied with a comparatively simple computational scheme which did not take up an excessive amount of computer time or storage. The first model used for prediction from real data (Charney et alii[2]) made the further assumption that the atmosphere was barotropic, and produced surprisingly good results, the derived motions closely resembling the flow at 500 millibars. In later models the three-dimensional, baroclinic structure of the atmosphere was taken into account by applying the vorticity equation at two or three levels in the atmosphere, enabling predictions to be made of the pressure pattern near the surface and of the vertical motion on a broad scale, giving some idea of the likely areas of cloud and precipitation. Within these models some allowance can be made for topography, friction and diabatic effects, and they can be made to provide useful indications of movement and development for periods up to 72 hours.

The main disadvantage of the filtered models is that they deal satisfactorily only with the major synoptic systems, on a scale of 1000 kilometres or more. Since many of the important weather-producing features are smaller than this, there are considerable benefits to be gained by the use of the primitive-equation (PE) model; it has the additional advantage that it can more readily and more directly incorporate the effects of friction, topography, turbulence, condensation, etc. The operational use of PE models has been made possible by the remarkable increases in speed and capacity of electronic computers in recent years, which made feasible the use of sufficiently short time-steps to overcome the problem of short-period oscillations, and, more recently, by improvements in the integration schemes which enable longer time-steps to be used than hitherto. PE models are in routine use in many countries, including the United Kingdom; the rest of this chapter will deal almost solely with PE models, and in particular with the 10-level model in use in Britain. The treatment will be, as far as possible, non-mathematical, although equations will sometimes be necessary to illustrate the basic principles. Nor will a great deal of detail be given because improvements are continually being made to the model, and it has been thought better to concentrate on aspects which are unlikely to change radically for some considerable time.

3.1.3 <u>The operational numerical analysis and forecasting system</u>

The forecast model is only one part of the overall numerical forecasting system. This starts with the extraction of the data required for the analysis from the coded messages received at the meteorological centre; wherever possible, quality control is carried out in various ways, such as checking against climatological values, ensuring that variations in time and space are reasonable and that there is internal consistency of the data within each message. Next is the analysis stage, in which polynomial functions are fitted to the data, both in the horizontal and in the vertical, to achieve a high degree of consistency. The data being analysed are in fact used to modify 'background' fields, forecasts for the appropriate time based on earlier data, so that an element of continuity is maintained and gaps in the observational material are filled. At this stage human intervention plays an important part, 'correcting' what may appear to be wrong decisions by the complex but rigid quality-control procedures (the human analyst is not always right), and amending the analyses either directly or by the insertion of 'bogus' observations. The topic of intervention is discussed in rather more detail in Chapter 11 - Upper-air charts.

In the filtered models, the analysed fields were completely specified by the distribution of geopotential on a given isobaric surface. For the primitive-equation models, however, it is necessary to define both the geopotential field and the wind-velocity field. The two fields must balance, otherwise fictitious motions are set up and amplify rapidly: the process of fitting the data in this way is known as 'initialization'.

After the analysis and initialization procedures have been completed, the forecast model itself takes over. The equations cannot be integrated analytically, and the calculations must be performed using the finite-difference forms. These mathematical techniques necessarily involve some approximation, which leads at times to errors, such as the slow movement of ridges and troughs in the westerlies; these can be recognized and reduced to some extent but not eliminated entirely. Physical processes, such as surface friction, diffusion, phase changes and precipitation of water, and exchange of heat with the surface, are included and allowance is made for the effects of topography and of small-scale convection.

Finally, the output must be put in a form to suit the customer. For some purposes, such as flight-planning, the output may be supplied to the customer without modification, while for others subjective amendment may at times seem worth while, as for example when forecasting upper winds for navigational purposes. Some forms of output are provided for the guidance of the subjective forecaster: the objective rainfall predictions and thermal vorticity charts, for example, are valuable for the estimation of frontal positions on the surface prognosis. More will be said about these topics in Chapter 13 - Computer prognoses: types and uses.

3.2 SPECIFICATION OF THE UNITED KINGDOM 10-LEVEL MODEL

The 10-level primitive-equation model actually comprises two models, the 'octagon' version and the 'rectangle' version. The octagon model covers the northern hemisphere north of about latitude 15°N; the grid points, just over 3000 in all, are regularly spaced on a polar stereographic map projection, with a separation of 300 km at 60°N. The rectangle model covers western Europe and much of the North Atlantic Ocean, the grid spacing being 100 km at 60°N on a polar stereographic projection, the total number of grid points again being rather over 3000. The grid systems for the two models are shown in Figure 1.

Pressure is used as the vertical co-ordinate, the atmosphere from 1000 mb to 100 mb being modelled at ten levels at intervals of 100 mb. The dependent variables, calculated from the prognostic equations (see 3.3 below) are the horizontal wind components (u and v), the thicknesses (h') of the 100-mb layers, and the geopotential of the 1000-mb surface (h_{10}). The geopotential of the other co-ordinate surfaces are derived from h_{10} and the values of h', while the vertical velocity is obtained from the equations of continuity. The humidity mixing ratio is also predicted in the layers up to 300 mb, but the atmosphere is assumed dry above.

In any finite model, the boundary conditions must be specified. In the octagon forecasts the lateral boundary values are assumed to remain constant throughout the period, while those for the rectangle model are derived from the octagon forecasts. The upper boundary condition for both models is that the vertical velocity is zero at 50 mb.

The following sections will deal with the remaining features of the model - the system of equations and methods of solution, the problems of analysis and initialization, and the additional effects incorporated in the model (e.g. topography, friction). Finally, the place of the model in the operational

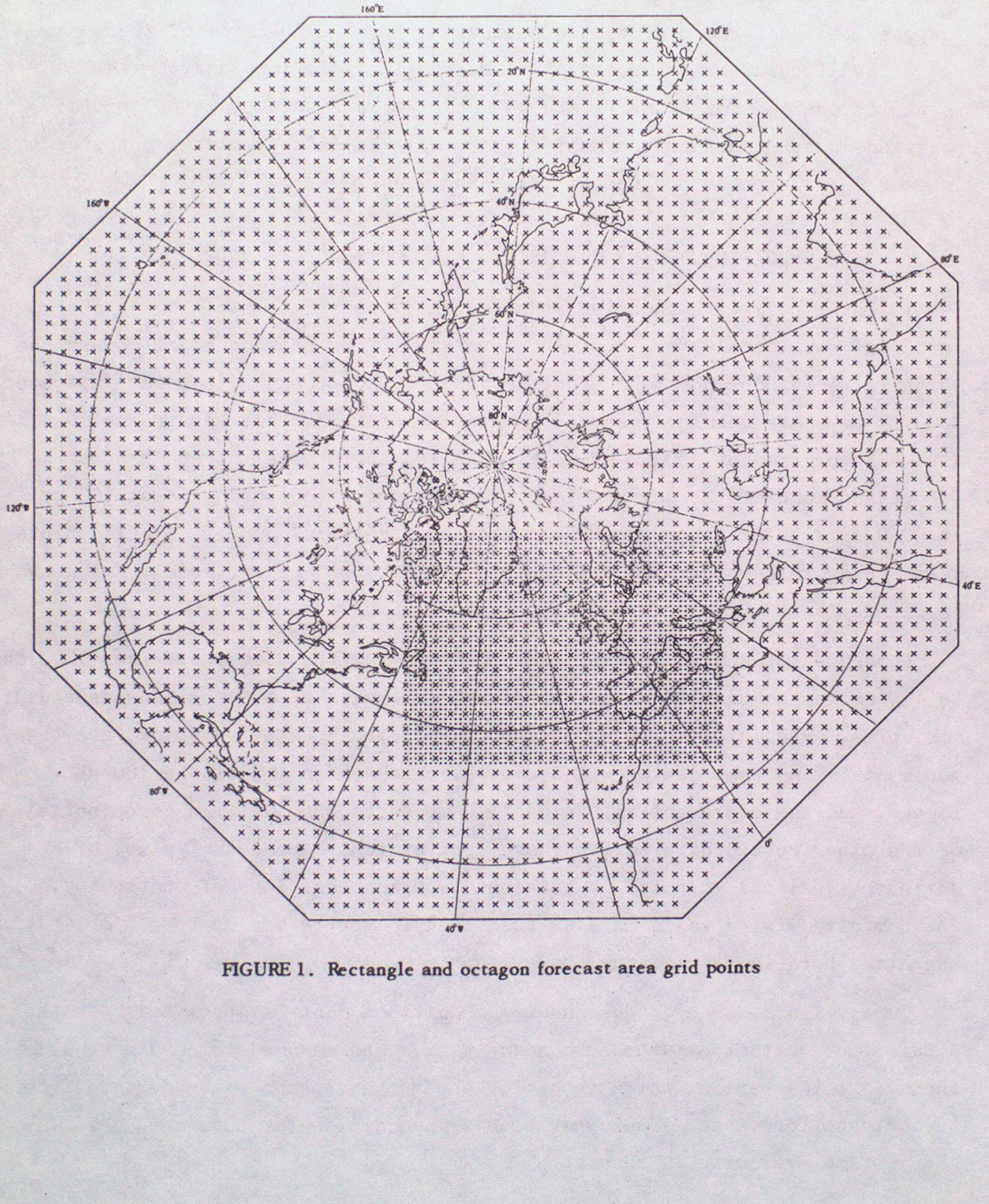


FIGURE 1. Rectangle and octagon forecast area grid points

forecast suite will be discussed, together with the methods of providing output for various requirements.

3.3 SYSTEM OF EQUATIONS

Since the horizontal co-ordinates are based on a polar stereographic map projection, the fundamental equations must be rewritten to allow for the variations of scale with latitude. This is done by the introduction of a map magnification factor, m, defined as the ratio of an infinitesimal distance on the map to the corresponding distance on the earth's surface: it is given by $m = 2/(1 + \sin\phi)$ where ϕ is the latitude.

The notation used is mostly standard, but for ease of reference a list is given at the beginning of this chapter (page ii). The treatment will be fairly brief; for greater detail the reader should consult References 3 and 4.

The equations may be simplified if we write:

$$\mathbf{V}^* = \mathbf{V}/m, \ \boldsymbol{\tau}^* = \boldsymbol{\tau}/m, \ \mu = m^2.$$

The vector operator, ∇, is two-dimensional, i.e.

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y}.$$

The principal equations comprise the momentum, thermodynamic and humidity equations and the equation of continuity. The first three are termed prognostic equations because they contain derivatives with respect to time. The last is purely diagnostic, relating the change of vertical velocity with pressure to the horizontal divergence of the flow. The hydrostatic equation, also diagnostic, is implicit in the derivation of the thermodynamic equation.

The basic equations are as follows:

(a) <u>Momentum equation</u>

$$\frac{\partial \mathbf{V}^*}{\partial t} + \mu(\mathbf{V}^*.\nabla)\mathbf{V}^* + \omega\frac{\partial \mathbf{V}^*}{\partial p} + f\mathbf{k}\wedge\mathbf{V}^* + g\nabla h + \tfrac{1}{2}|\mathbf{V}^*|^2\,\nabla\mu = -g\frac{\partial \boldsymbol{\tau}^*}{\partial p} \quad \ldots\ldots (3.1)$$

The first three terms of this equation add up to the horizontal component of the acceleration experienced by a given particle embedded in the flow, that is the acceleration following the flow. (A more familiar form of equation for the total derivative, $\mathrm{d}/\mathrm{d}t$, (the rate of change of a property following the flow) is

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}$$

where $\partial/\partial t$ is the local rate of change of the property, and $w(\partial/\partial z) = \omega(\partial/\partial p)$.) The fourth term is the Coriolis term, $\mathbf{k}$ being a unit vertical vector, while the fifth term is the pressure gradient force. The final term on the left-hand side (l.h.s.) is to compensate for changes of the map magnification factor. The right-hand side (r.h.s.) of the equation represents the effect of the surface frictional stress. Equation 3.1 is simply a statement of Newton's second law of motion, which postulates that 'the rate of change of momentum of a body, relative to axes fixed in space, is equal to the vector sum of the forces acting on that body'.

(b) Thermodynamic equation

$$\frac{\partial^2 h}{\partial p \partial t} + \mu \mathbf{V}^*.\nabla\left(\frac{\partial h}{\partial p}\right) + \omega \frac{\partial^2 h}{\partial p^2} + \frac{\omega(1-\kappa)}{p} . \frac{\partial h}{\partial p} = - \frac{\kappa Q}{g p} \qquad \ldots . (3.2)$$

In this equation, h is the height of **an isobaric surface**, and $\partial h / \partial p$ is the thickness of a layer of unit pressure difference. The equation may seem complex, but its physical meaning is straightforward enough. The last term on the l.h.s. and the term on the r.h.s. together relate the added heat, Q, to the changes of state of the air, provided that the state of the air is not influenced by its motion. The first three terms on the l.h.s. give the rate of change of thickness, following the flow, of a layer of unit pressure difference: through the relationship between the thickness of a layer and its mean temperature they may be regarded as representing temperature changes.

Alternatively, the third and fourth terms can be shown, after some manipulation and substitution of the hydrostatic relationship, to be proportional to the product of the vertical velocity, ω ($= dp/dt$), and the difference between the actual lapse rate, $-(\partial T/\partial z)$, and the dry-adiabatic lapse rate, Γ_d, that is to say they provide a measure of the static stability. The local thickness changes (first term on l.h.s.) can then be seen to be the result of horizontal advection (second term on l.h.s.), vertical motion (third and fourth terms on l.h.s.) and added heat (r.h.s.).

If we replace $(\partial h/\partial p)$ by $(-h'/\Delta p)$, where h' is the thickness of a layer defined by the pressure difference Δp, equation 3.2 becomes

$$\frac{\partial h'}{\partial t} + \mu \mathbf{V}^*. \nabla h' + \omega \beta = \frac{\kappa Q \Delta p}{g p} \qquad \ldots . (3.2a)$$

where $\beta = \left[(\partial h'/\partial p) + \omega(1-\kappa)/p\right]$, the 'stability term'. It is this form of the equation which is used, with Δp = 100 mb, in the 10-level model, h' now being the thermodynamic variable.

(c) Humidity equation

$$\frac{\partial r}{\partial t} + \mu \mathbf{V}^*. \nabla r + \omega \frac{\partial r}{\partial p} = W. \qquad \ldots . (3.3)$$

This equation relates the change of humidity mixing ratio, following the flow, to the amount of water vapour added or lost by evaporation, condensation or precipitation.

(d) Continuity equation

$$\frac{\partial \omega}{\partial p} + \mu \operatorname{div} \mathbf{V}^* = 0. \qquad \ldots . (3.4)$$

The vertical velocity, ω, is calculated by integrating $\partial\omega/\partial p$ from the 50-mb surface downwards using the upper boundary condition, $\omega = 0$ at 50 mb.

3.4 FINITE-DIFFERENCE SCHEMES

3.4.1 Introduction

Whereas the basic physics of the model as so far described are straightforward enough, the development of satisfactory and efficient means of integrating the differential equations poses many problems, and a good deal of effort has had to be expended in finding solutions.

The integration process requires the independent variables to be specified at a set of grid points: this is the function of the analysis and initialization stages which will be described in the next section. Finite-difference estimates of the space derivatives of the independent variables (h', u^*, v^* and r) are found for each of the appropriate grid points, and the finite-difference forms of the governing equations are used to obtain the time derivatives. These are then used in a truncated Taylor series to yield the new values of the variables one time-step, δt, ahead. Not all the variables are specified at all points of the grid, the precise disposition of the variables, and hence the precise form of the finite-difference equations, depending upon the integration scheme being used. Three such schemes have been used in the four years since the introduction of the 10-level model into operational use in August 1972. A full description of these schemes is well outside the scope of this handbook, but an attempt will be made to outline the principles on which the calculations are based; for further details the reader should consult References 3 and 4.

3.4.2 The Lax-Wendroff scheme

The basis of the Lax-Wendroff scheme is the calculation of the value (Z_{n+1}) of a variable Z at time $(n+1)\delta t$ from its value, Z_n, and the first two time derivatives $\dot{Z}_n = (\partial Z/\partial t)_n$, $\ddot{Z}_n = (\partial^2 Z/\partial t^2)_n$ at time $n\,\delta t$, using a truncated Taylor series:

$$Z_{n+1} = Z_n + \delta t\,\dot{Z}_n + \tfrac{1}{2}\delta t^2\,\ddot{Z}_n \qquad \ldots . \, (3.5)$$

The time derivatives cannot be calculated directly and a two-step approximation has to be used. The first step is a forward time-step over half the period of integration:

$$Z_{n+\frac{1}{2}} = Z_n + \tfrac{1}{2}\delta t\,\dot{Z} \qquad \ldots . \, (3.6a)$$

New values of the variables at $t = (n + \frac{1}{2})\delta t$ are then calculated and used in the second step, which is a centred time-step over the whole of the integration period, δt:

$$Z_{n+1} = Z_n + \delta t\,\dot{Z}_{n+\frac{1}{2}} \qquad \ldots . \, (3.6b)$$

By differentiating equation 3.6a with respect to time and substituting in equation 3.6b, the process can be seen to be equivalent to the use of equation 3.5. The time derivatives, $\dot{Z}_n$ and $\ddot{Z}_n$, are calculated, using the finite-difference form of the primitive equations by the methods indicated below.

The 'staggered grid' used with the Lax-Wendroff integration is shown in Figure 2, with the disposition of the variables at times $n\,\delta t$, $(n+\frac{1}{2})\delta t$ and $(n+1)\delta t$. It will be noted that there is only one variable at most grid points, and that the grid for the intermediate time is displaced half a grid length in each of the x- and y-directions from that at the main times. This 'staggering' is essential for the proper operation of the Lax-Wendroff scheme. The vertical disposition of the variables is shown in Figure 3; u, v and h are calculated at the ten isobaric co-ordinate surfaces, while h' (or $\partial h/\partial p$), r and ω are derived at the intermediate levels, with ω also at 1000 mb.

The use of the staggered grid and the finite-difference form of the equations will now be illustrated by a simple example. Figure 4(a) shows a small section of the grid with the variables appropriate to each grid point. In the calculation of a new value of v at A at $(n+1)\delta t$, it is first of all necessary to calculate the variables at the intermediate time, $(n+\frac{1}{2})\delta t$, as shown in Figure 4(b). It will be noted that the variable u occurs at A at the intermediate time, so that one of the values required is $u_{n+\frac{1}{2}}$ at A at time $(n+\frac{1}{2})\delta t$: this is calculated from the momentum equation (page 7) and equation 3.6a. Consider the advection term for the u-component of the momentum equation:

$$u\,\frac{\partial u}{\partial x} + v\,\frac{\partial u}{\partial y}\,;$$

this becomes, in terms of the averages and finite differences between grid points,

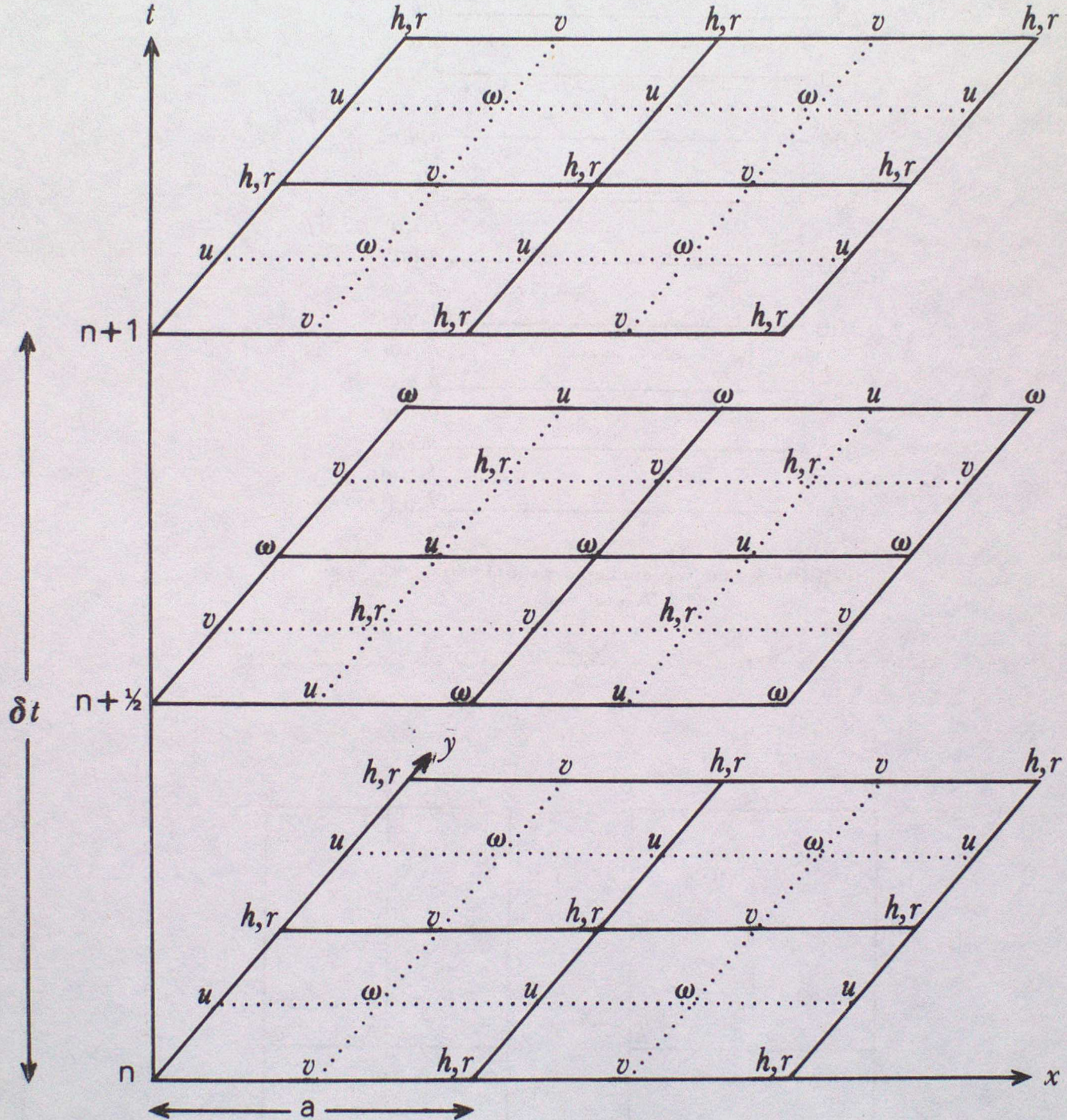


FIGURE 2. Staggered grid used in the Lax-Wendroff integration scheme

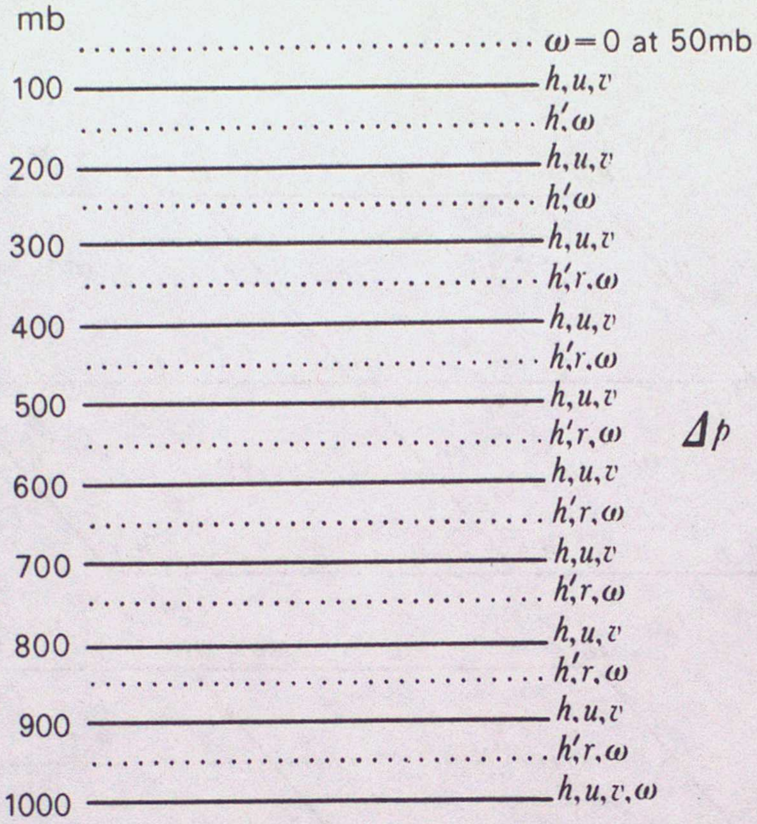
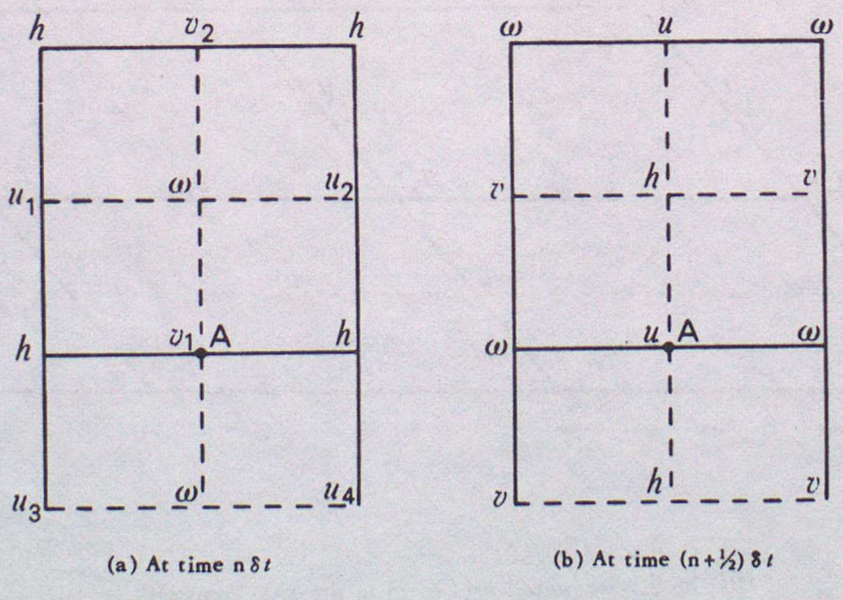


FIGURE 3. The vertical deployment of variables in the 10-level model



(a) At time $n\delta t$

(b) At time $(n+\frac{1}{2})\delta t$

FIGURE 4. Diagram to illustrate the finite-difference scheme

$$\tfrac{1}{4}(u_1+u_2+u_3+u_4)\times\frac{1}{2a}\Big[(u_2+u_4)-(u_1+u_3)\Big]+\frac{v_1}{2a}\Big[(u_1+u_2)-(u_3+u_4)\Big].$$

The remaining terms of the equation are calculated in a similar way, the time derivative of u, $\dot{u}$, is calculated from the momentum equation and used to compute $u_{n+½}$. The other grid-point variables are treated in the same way, and the whole process repeated to obtain the values of the variables at $(n+1)\delta t$.

Before the vertical velocity is computed from the continuity equation, the effects of frictional stress and horizontal diffusion are incorporated. This is done by means of a simple forward step over the whole of δt, the product of δt and the tendencies, $\dot{Z}_F$ and $\dot{Z}_D$, due to these effects simply being added to equation 3.6b:

$$Z_{n+1} = Z_n + \delta t\,\dot{Z}_{n+½} + \delta t\,(\dot{Z}_F)_n + \delta t\,(\dot{Z}_D)_n$$

where $\dot{Z}_D$ is the finite-difference equivalent of $K\nabla^2 Z = \Big[(\partial^2 Z/\partial x^2)+(\partial^2 Z/\partial y^2)\Big]K$, and Z is the meteorological quantity under consideration, e.g. u or v. This term, being a second-order derivative, has to be calculated from grid-point values over a wider area than is necessary for the first space derivatives. The evaluation of $\dot{Z}_F$ is discussed in 3.6.2.1 (page 21). Then, remembering that ω is found at intermediate levels, we have for the calculation of ω at one level $(k+½)$ from its value at another level $(k-½)$,

$$\omega_{k+½} = \omega_{k-½} - \frac{\Delta p.\mu}{a}\Big[(u_2-u_1)+(v_2-v_1)\Big],$$

the finite-difference form of the continuity equation.

Finally, the effects of Q and W are added to the values of b'_{n+1} and r_{n+1} as outlined in 3.6.2.2 (page 21).

3.4.3 The split semi-implicit integration scheme

The main disadvantage of most explicit integration schemes for use with primitive-equation models is that they suffer from instability unless the time-step is very short. The longest permissible time-step depends upon the grid length and wind speed; for the rectangle model, with a = 100 km, the maximum time-step is about three minutes. This time is determined by the speed of fast-moving gravity waves which, although forming only an insignificant part of the pressure variation of synoptic-scale systems, may amplify unrealistically in the model and disrupt the calculations. Implicit schemes do not have the same stringent stability criterion, and appreciably longer time-steps may be used to speed up the model's calculations, but there is a disadvantage in that more complex arithmetical operations are necessary. Why this is so can be seen if we consider the nature of the implicit integration scheme. For the equation

$$\frac{\partial X}{\partial t} = F(X)$$

an explicit integration scheme might be

$$X_{n+1} = X_{n-1} + 2\delta t\, F(X_n),$$

defining X_{n+1} explicitly in terms of values at earlier times. On the other hand, the integration represented by

$$X_{n+1} = X_n + \frac{\delta t}{2}\left[F(X_n) + F(X_{n+1})\right]$$

defines X_{n+1} implicitly in terms of itself and earlier values, and the equation must be solved by matrix inversion, often by an iterative method. A compromise is attained in the semi-implicit schemes in which some terms are treated explicitly, while those which govern the motion of gravity waves are treated implicitly. The gravity waves have different modes which travel at different speeds, and in the scheme devised by Burridge,[5] and used for some time in the 10-level model, it is necessary to treat only the two fastest of ten possible gravity waves implicitly, giving a considerable saving of computation time.

The treatment of the equations is split into two stages - the 'advective' and the 'adjustment' stages. For the momentum equation (3.1 on page 7), for example, the terms are split as follows:

(a) <u>Advective stage</u>

$$\frac{\partial \mathbf{V}^*}{\partial t} = -\mu(\mathbf{V}^*.\nabla)\mathbf{V}^* - \omega\frac{\partial \mathbf{V}^*}{\partial p} - \frac{1}{2}|\mathbf{V}^*|^2\nabla\mu - g\frac{\partial \tau^*}{\partial p} + mK\nabla^2(m\mathbf{V}^*)$$

(the last term being added now to allow for lateral diffusion). This is integrated forward in time through a single time-step, δt, using the Lax-Wendroff explicit scheme described in the preceding section.

(b) <u>Adjustment stage</u>

$$\frac{\partial \mathbf{V}^*}{\partial t} = -f\mathbf{k}\wedge\mathbf{V}^* - g\nabla b.$$

After some manipulation, the equations for the adjustment stage yield solutions which correspond to ten gravity-wave modes, each characterized by a different propagation speed. The three fastest travel at speeds, c, of roughly 300, 100 and 50 m s^{-1}: the maximum time-step is given by

$$\Delta t = a/(\sqrt{2}c)$$

so that for $c = 50$ m s^{-1} a maximum time-step of rather over 20 minutes may be used and gravity waves of this and lower speeds may be treated explicitly. The two fastest must, however, be treated implicitly if such a time-step is to be used.

The introduction of the split semi-implicit scheme achieved substantial savings in computer time - by a factor of four for the rectangle forecasts and two for the octagon forecasts.

3.4.4 The split explicit scheme

Again the governing equations are split into advection and adjustment stages as in the split semi-implicit method. The advection stage is integrated by a modified Lax-Wendroff scheme in which the first space derivatives are evaluated to fourth-order accuracy (instead of the second order as in the original scheme). The adjustment stage is carried out in three short explicit steps instead of a single implicit step. The result is that a longer time-step of 30 minutes can be used for the advective stage, resulting in a reduction in computer time by a factor of two over the semi-implicit scheme. A more accurate treatment of the Coriolis term is possible than with the semi-implicit scheme, which at times led to spurious development of intense cold pools. The changes resulting from physical processes are now incorporated into the adjustment stage.

3.4.5 Lateral boundary conditions

In the octagon model the geopotentials of the isobaric surfaces, the humidity mixing ratios and wind velocities at the boundary are kept constant throughout the forecast period. For the rectangle forecasts the boundary conditions are mostly derived from the tendencies interpolated from the octagon model. In both models, stability near the boundary is maintained by using an enhanced horizontal diffusion coefficient in a narrow boundary zone.

3.5 ANALYSIS AND INITIALIZATION

Before the procedures of 3.4 can be put into operation, it is necessary to specify the initial values of the variables at the appropriate grid points and levels. The means whereby this is done can be divided into two stages - analysis and initialization.

The analysis stage produces fields, or sets of grid-point values, of the geopotentials of isobaric surfaces and of the components of the wind velocity. Initialization is designed to ensure that the wind and height fields are carefully balanced in order to avoid the initiation of fast-moving and rapidly amplifying gravity waves.

3.5.1 Analysis

The analysed field for a given time and pressure level is based upon the following:

(a) a background field, usually a 12-hour forecast based on the analyses for 12 hours earlier;

(b) the observations for that time;

(c) human intervention; and

(d) for certain non-standard levels, interpolation from the standard levels.

The background field is necessary to fill gaps in areas where observations are sparse or missing entirely, and also at times to check the validity of an observation. The method of analysis for contour height which was used when the 10-level model was first introduced was to fit a quadric surface to the observations in the vicinity of each grid point, the weight given to each observation decreasing with increasing distance from the grid point. The surface had the form

$$h_s = ax^2 + hxy + by^2 + 2gx + 2fy + c$$

where h_s is the geopotential of the quadric surface at the point (x, y), the origin of co-ordinates being at the grid point. The coefficients a, b, c, f, g, h are found by the method of least squares, and the procedure is repeated for every grid point. Allowance was also made for reported winds, to determine the local contour gradient near an observation and hence, in regions with sufficient observations, to provide new estimates of the geopotential at one or more grid points. Curvature of the flow was also taken into account (see Corby[6]). The precise way in which the computer model proceeds with the analysis to ensure vertical consistency and to allow for interpolation between the standard analysis levels is discussed in a little more detail in Chapter 11 - Upper-air charts, as also are the methods of intervention designed to improve the analysis at the main standard levels and, indirectly, at all levels.

The analysis of relative humidity has been described by Atkins:[7] it is carried out in much the same way except that the weighting factors fall off much more rapidly in a direction perpendicular to the isopleths, particularly when the gradient is strong, than they do in a direction parallel to the isopleths. Intervention is again used and, in particular, relative humidities may be estimated from a study of visible and infra-red satellite photographs of cloudiness.

An improved method of geopotential analysis, known as orthogonal polynomial analysis, is currently in use. The theory has been described by Dixon[8] and by Dixon and Spackman.[9] The practical application is similar to the grid-point analysis described above in so far as the analysis proceeds level by level, each level being scanned a number of times to allow data rejected by the quality-control checks to be excluded from the following scans. The geopotential at each grid point is derived as in the grid-point analysis (see Chapter 11 - Upper-air charts), using wind data to supplement the height information. No correction is made for the asynoptic nature of satellite and aircraft data, while the latter are also assigned to the nearest standard level without correction; however, non-standard data are given a lower weighting than synoptic data.

The next step is that which gives its name to the overall procedure: the entire set of grid-point geopotentials is fitted, in the least-squares sense, to a series expansion of empirical orthogonal polynomials. A full description is outside the scope of this handbook and of interest mainly to specialists, so only a very brief, simplified account will be given here.

In one dimension, for any set of points x_i, where $i = 1, 2, ..., N$, will exist a set of polynomials of the form

$$f_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \qquad \ldots . (3.7)$$

where $n = 0, 1, \ldots, (N-1)$, such that for all combinations of m and n

$$\sum_{i=1}^{N} f_m(x_i)\, f_n(x_i) \begin{cases} = 0 & \text{for } m \neq n. \\ \neq 0 & \text{for } m = n. \end{cases} \qquad \ldots . (3.8)$$

Such a set of polynomials is said to be orthogonal over the set of points, x_i. A more familiar interpretation of equation 3.8 is that $f_m(x_i)$ and $f_n(x_i)$ are not correlated when $m \neq n$, the left-hand side of the equation being analogous to the numerator in the expression for a correlation coefficient.

If b is a quantity which varies with x, and the values of b are known for all values of x_i, then b may be expressed in terms of the N polynomials: at each point

$$b_i = \alpha_0 f_0(x_i) + \alpha_1 f_1(x_i) + \ldots + \alpha_{N-1} f_{N-1}(x_i) \qquad \ldots . (3.9)$$

Equation 3.9 is a set of N simultaneous equations, from which the N coefficients, $\alpha_0 \ldots \alpha_{N-1}$, may be determined, in principle uniquely. In practice, the degree, q, of the highest-order polynomial to be fitted, is much less than N, and the number of equations consequently much greater than the number of unknowns: the 'best' values of the coefficients must therefore be found in some way. Usually this is done by the method of least squares, a basically similar process to that used for simple linear regression (see

Chapter 24 - Local forecast studies). In fact, the least-squares solution for the determination of the q + 1 coefficients, as will be seen in the Appendix (page 29), is particularly simple. The coefficients are given by expressions of the form

$$\alpha_n = \frac{\Sigma h f_n}{\Sigma f_n^2} \quad \ldots \ldots (3.10)$$

where the summation is taken over all values of x_i, i.e.

$$\Sigma h f_n = \sum_{i=1}^{N} h(x_i)\, f_n(x_i).$$

The method can readily be extended to two (or more) dimensions, which allows, say, a geopotential field to be expressed in terms of the horizontal co-ordinates, x and y, of a network of grid points:

$$h(x,y) = \sum_{i,j} \beta_{ij}\, \phi_{ij}(x,y)$$

where ϕ_{ij} is a set of polynomials orthogonal over the set of grid points. The least-squares solution for the coefficients, β_{ij}, is simply

$$\beta_{ij} = \frac{\Sigma h \phi_{ij}}{\Sigma \phi_{ij}^2}$$

where the summation is taken over the network of points.

In practice, in the objective-analysis model, since the grid is square, $\phi_{ij}(x,y)$ may be reduced to $\phi_i(x)\ \phi_j(y)$: if this could not be done the method would at present be impracticable because of the enormous demand on the computer. The power, p, of the highest-order polynomial is much less than the number of grid points, N, in a row or column of the grid. The fitted heights are given by the equation

$$h_p(x,y) = \sum_{i,j=0}^{p} \beta_{ij}\, \phi_i(x)\ \phi_j(y)$$

where p must be < N. The value of p used in the model varies with the pressure level and scan, with a maximum value of over 30.

Any observation which differs by too great an amount from a fitted surface can readily be spotted, and amended or rejected if necessary, providing a valuable addition to the quality-control procedures.

When all pressure levels have been analysed, the coefficients β_{ij} are fitted in the vertical by another set of orthogonal polynomials, giving a three-dimensional fit to the original data and helping to ensure vertical consistency in the analysis.

3.5.2 Initialization

In general, the primitive equations are capable of predicting the motion of gravity waves. Gravity waves of small amplitude occur in the real atmosphere, but when analysing initial conditions it is not easy to distinguish between those wind and height fields which lead to realistic gravity-wave predictions, and slightly different fields which lead to the production of large-amplitude gravity waves which disrupt the forecast calculations. For this reason, the initialization stage has been designed to eliminate gravity waves entirely during the initial stages of the forecast.

Thompson[10] showed that a rotating atmosphere in hydrostatic equilibrium is unable to support gravity-wave propagation if the total rate of change of the divergence, D, of the horizontal wind is zero, i.e.

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \left[\frac{\partial}{\partial t} + m\left(u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right) + \omega\frac{\partial}{\partial p}\right]D = 0 \qquad \ldots. (3.11)$$

and, to ensure that this condition does not change with time,

$$\frac{\partial}{\partial t}\left(\frac{\mathrm{d}D}{\mathrm{d}t}\right) = 0. \qquad \ldots. (3.12)$$

These equations, however, are not in a suitable form for an analysis scheme, since they contain time derivatives, and they must be converted into purely diagnostic expressions for use in the initialization procedure. This is done for equation 3.11 by substituting for $\partial u/\partial t$ and $\partial v/\partial t$ from the momentum equation (page 7); ignoring friction and diffusion we now have a form of the balance equation:

$$m\left[\left(\frac{\partial u}{\partial x}\right)^2 + 2\left(\frac{\partial v}{\partial x}\right)\left(\frac{\partial u}{\partial y}\right) + \left(\frac{\partial v}{\partial y}\right)^2\right] - f\zeta + mg\nabla^2 h + \mathbf{k}\wedge\mathbf{V}\cdot\nabla f + g\nabla m\cdot\nabla h = 0. \qquad \ldots. (3.13)$$

To fulfil the condition of equation 3.12, the local derivative of equation 3.13 with respect to time should also equal zero: differentiating and taking only the most important terms gives:

$$\frac{\partial}{\partial t}(-f\zeta + mg\nabla^2 h) = 0 \qquad \ldots. (3.14)$$

Substituting for $\partial\zeta/\partial t$ from the vorticity equation and for $\partial h/\partial t$ from the thermodynamic equation, a form of the omega-equation is obtained,

$$m^2\nabla^2(\sigma\omega) + f(m\zeta + f)\frac{\partial^2\omega}{\partial p^2} + gm^2\nabla^2\left(m\mathbf{V}\cdot\nabla\frac{\partial h}{\partial p}\right) - fm\frac{\partial}{\partial p}\left[\mathbf{V}\cdot\nabla(m\zeta + f)\right] = 0 \qquad \ldots. (3.15)$$

where $\sigma = -1/\rho\,[\partial(\ln\theta)/\partial p]$.

In principle, equations 3.13 and 3.15, together with the continuity equation,

$$m\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{\partial\omega}{\partial p} = 0 \qquad \ldots. (3.16)$$

may be solved simultaneously to derive the fields of h, u, v and ω, but the amount of computation is formidable, and in practice much shorter methods of solution can be found (see Reference 3).

Having undergone the analysis and initialization procedures, the 'actual' fields are now ready to be used as input for the forecast scheme. The initialization procedure, and to a lesser extent the analysis procedure, will have introduced a degree of over-smoothing which may affect the fields, particularly of vertical velocity, for the first few hours of forecast time, but which has little effect on the forecast fields which are actually used - for 12 hours and longer periods ahead.

3.6 ADDITIONAL EFFECTS

The 10-level model takes into account the effects of a number of physical processes through the terms Q, W, τ and K: these include the addition and removal of heat, changes of phase of water and precipitation, surface friction and lateral diffusion. The effects of topography are included, the height of the ground being a substantially smoothed representation of the actual surface.

3.6.1 Water: phase changes and precipitation

The phase changes of water and the precipitation of water from the atmosphere contribute to the terms Q and W in the basic equations. Each of the 100-mb layers of the model is tested to see if the actual humidity mixing ratio, r, is greater than the saturated mixing ratio, r_s, (with respect to water or ice as appropriate), the values of r_s being derived from polynomial expressions in terms of the layer thicknesses. If $r_s > r$, condensation of the excess occurs and latent heat is released.

In the model there is no storage of liquid water in any layer: any liquid water or snow falls through the layers below. Evaporation occurs if the precipitation falls through an unsaturated layer, the latent heat of evaporation used resulting in a cooling of the layer. Any such evaporation and cooling that has occurred in previous steps is taken into account when testing the layer for supersaturation. Liquid water or snow reaching the ground is interpreted as surface precipitation. The melting layer is that which contains the (highest) $0^{\circ}C$ isotherm; in layers above this any precipitation is taken to be snow, and any below this layer to be rain. If the temperature is above $-5^{\circ}C$, precipitation is assumed to form at a rate sufficient to keep the air just saturated with respect to liquid water. Below $-5^{\circ}C$, snowflakes are grown at a temperature intermediate between water and ice saturation (see Wexler[11]). The latent heat of melting is taken into account when calculating the thickness of the melting layer and testing it for supersaturation.

3.6.2 Surface exchanges

3.6.2.1 Surface friction. Surface friction is a manifestation of the exchange of momentum between the lowest layers of the atmosphere and the underlying surface: it is represented in the equations of motion by $-g(\partial \boldsymbol{\tau}/\partial p)$. If $\boldsymbol{\tau}_{10}$ is the surface (1000-mb) stress vector and $\mathbf{V}_{10}$ the 1000-mb wind velocity, then

$$\boldsymbol{\tau}_{10} = \rho_{10} C_D |\mathbf{V}_{10}| \mathbf{V}_{10} \qquad \ldots . (3.17)$$

where ρ_{10} is the density of air at 1000 mb in the International Standard Atmosphere and C_D is the surface drag coefficient. Values of C_D used in the model are

(a) over sea: $C_D = 1.0 \times 10^{-3}$

(b) over land: $C_D = \left(1.0 + \frac{H}{400}\right) \times 10^{-3}$

where H is the smoothed topography height in metres. The variation with H over land is introduced to allow for excess drag over high ground without introducing a large discontinuity at a coastline.

In the model it is assumed that the friction layer is confined to the lowest model layer, i.e. at 900 mb $\boldsymbol{\tau} = 0$ and $\partial \boldsymbol{\tau}/\partial p = 0$. The profile is assumed to be quadratic, so that

$$\left(\frac{\partial \boldsymbol{\tau}}{\partial p}\right)_{10} = \frac{2\boldsymbol{\tau}_{10}}{\Delta p} . \qquad \ldots . (3.18)$$

The value of $(\partial \boldsymbol{\tau}/\partial p)_{10}$ is then used in equation 3.1 (page 7) to calculate the time derivatives, $\dot{u}_F^*$ and $\dot{v}_F^*$, of the 1000-mb wind-velocity components resulting from frictional losses of momentum.

3.6.2.2 Surface exchanges of sensible and latent heat. Formulae similar to equation 3.17 are used to calculate the flux of sensible heat, F_H, across the 1000-mb surface, the net flux depending upon the difference between the values of temperature or mixing ratio at the surface and at a level in the atmosphere close to the surface (the 10-metre level is generally used). The formulae are:

$$\left.\begin{aligned} F_H &= \rho_{10} C_E |\mathbf{V}_{10}| c_p (T_o - T_a) \\ F_W &= \rho_{10} C_E |\mathbf{V}_{10}| (r_o - r_a) \end{aligned}\right\} \qquad \ldots . (3.19)$$

where C_E is analogous to the drag coefficient.

T_a and r_a are obtained from the 10-level model on the following assumptions:

(a) if the air is heated from below, there is a dry-adiabatic lapse rate and the relative humidity is constant from 10 metres to 950 mb;

(b) if the surface is colder than the air, $(T_o - T_a) = 0.1(T_o - T_a^1)$, where T_a^1 is the value obtained for the near-surface temperature by applying an extrapolation of the 900-mb lapse rate to the 950-mb temperature. One tenth of the relative humidity change from the surface to 950 mb is assumed to take place in the lowest 10 metres.

Over the sea, the surface temperature T_o is assumed to remain constant over the forecast period, and the humidity mixing ratio is taken as its saturation value. The exchange coefficient is given by

$$C_E = \left[1 + 0.1\,(T_o - T_a)\right] \times 10^{-3}$$

taking some account of the variation of C_E with stability.

Over land, the fluctuations of T_o are generally so great that equations 3.19 cannot be used, and the calculation of F_H and F_W is more complex, taking into account the net radiation flux at the surface, R_N, the flux of heat into the ground, F_G, and the partitioning of the available energy into sensible and latent heat. The net radiation flux depends upon solar altitude, calculated by the use of standard astronomical formulae, and cloudiness, estimated from the model's relative humidities by the method described by Ricketts.[12]

The heat flux into the ground, F_G, is fairly small, and is given with sufficient accuracy by XR_N, where

$$\begin{aligned} X &= 0.0 \text{ for snow-covered surfaces,} \\ \text{otherwise } X &= 0.1 \text{ for } R_N > 0, \\ \text{or } X &= 0.5 \text{ for } R_N \leqslant 0. \end{aligned}$$

For $R_N \leqslant 0$ it is assumed that $F_W = 0$ (neglecting dew formation) and $F_H = 0.5\,R_N$. For $R_N > 0$, a Penman-type method is used to determine the amount of evaporation and the heat required to bring this about. The rest of the available energy appears then as sensible heat.

Surface energy fluxes over sea ice are also included, and allowance is made for the heat used in melting sea ice where appropriate.

3.6.3 Convective adjustments

The effects of convection on scales smaller than that of the grid-point network must be allowed for to prevent the growth in the model of unrealistic convection on the scale of the grid itself. Convective adjustment schemes designed for this purpose also allow for the redistribution of heat and moisture throughout the lower atmosphere by convection.

It is assumed that the vertical transport of heat by sub-grid-scale convection maintains a neutral lapse rate on the grid scale whenever this would otherwise be exceeded. The neutral lapse rate for the purposes of this model varies between the dry-adiabatic lapse rate, when the relative humidity is less than a critical value, to the saturated-adiabatic lapse rate for saturated air. It is further assumed that vertical transport of water vapour leaves the grid-scale relative humidity unchanged. Excess water vapour is allowed to condense: if it reaches the surface it is regarded as convective rainfall. The final assumption is that the vertically integrated sum of potential, internal and latent energies is conserved, i.e.

$$\int_{p=1000}^{100} (c_p T + L r)\,dp = \text{constant.}$$

The mechanics of the operation is first of all, after each time-step, to add the thickness increments $\delta h'$ from surface fluxes to the thickness, $h'_{9.5}$, of the lowest layer. The corresponding humidity increment is added to those layers involved in the convective adjustment at the end of the previous time-step. A check is made to see if supersaturation occurs in any layer affected; if it does, the surface flux of water vapour is reduced to prevent this happening. The convective adjustment is then carried out as indicated earlier.

Convective adjustments are also made in the analysis stage - otherwise there might be a sudden change in the lapse rate or spurious convective rainfall in the early stages of the forecast model.

The scheme outlined above allows only for shallow convection, but a more comprehensive model, which copes with deep convection, has recently been introduced into the operational routine.

3.7 THE OPERATIONAL SUITE

The forecast model itself is just one part of the 'operational suite' of programs which deal with every process from the receipt of the data to the despatch of the forecast products to the users.

Incoming data are first checked for correct format and then put into the Synoptic Data Bank. Quality control and further checking procedures (see Chapter 11 - Upper-air charts) are carried out, and suspect data may be rejected or corrected by human intervention before the data required for the analysis are selected and stored in the Basic Analysis Data Sets.

Data come in for many hours after the main synoptic observation time, rapidly at first and then at a rate which decreases with time. A compromise has to be reached between the opposing needs for timeliness, i.e. for the analyses to be available as soon as possible after the observation time, and for accuracy, i.e. for the analyses to be based on as complete a data set as possible. In current practice, two runs* are carried out:

(a) the 'operational' run, with a 'cut-off time' of HH + 0220 for the rectangle run and HH + 0315 for the octagon run; and

(b) an 'update' run, with a cut-off time of HH + 1100.

The operational run, as its name implies, produces the model's operational analyses and forecasts. There may be further human intervention, based on the results of these analyses, on any of the incoming data. 'Bogus' data may also be inserted to amend the analyses in certain areas, and direct modification of the background fields may be carried out by means of a visual display unit. The update run is then carried out, its main purpose being to provide the best possible set of analyses and a 12-hour forecast for use as a background field in the next operational run.

In the output phase, the forecast data are put into a common data set. Some processing is carried out during this stage: for example, the wind components and heights are held at different points in the grid, but the output requires wind speed and direction and height at the same points. Other, indirect output may also be derived, such as the maximum wind and the height at which it occurs, the vertical temperature profile and the level of the tropopause, and so on. Once this processing has been done, the various output programs may access the data set. This system has the advantage that modifications can be made to the format of the data from the forecast model, such as occurred when the integration scheme was changed, without any changes being necessary in the output programs.

Output may be in the form of paper or magnetic tape (for transfer to another computer) or in chart form for visual inspection by the forecaster. The types of output available are discussed in Chapter 13 - Computer prognoses: types and uses.

3.8 CONCLUDING REMARKS

Rapid strides have been made in the past decade in the speed and power of electronic computers, and advances in meteorological numerical models have at least kept pace with, and have often anticipated, computer technology. Nevertheless, there is still a great demand for the skill and judgement of the human forecaster in ensuring that the quality-control and analysis stages

---

*An intermediate run has recently been introduced to provide analyses of data received by HH + 0700 as an aid in the intervention process.

produce the best possible results. There is equally a need for the human forecaster in the use and interpretation of the numerical forecasts. No numerical model is perfect, errors arising both in defining the initial state (i.e. arising from the data, analysis and initialization) and in the forecast stages from the finite and discrete nature of the grid and in the integration of the differential equations. The human being, however, can study the output of the model in a variety of situations, and make allowances for known error tendencies when preparing a subjective forecast or modifying an objective one for some purpose. Even though for many uses the objective prognoses, particularly the upper-air ones, are satisfactory as they come from the computer, there seems little likelihood that the human component of the 'man-machine mix' will become redundant. Indeed, with increasing complexity of the models there may even be a greater demand for subjective intervention of some kind.

BIBLIOGRAPHY

1. RICHARDSON, L.F.; Weather prediction by numerical process. Cambridge University Press, 1922.

2. CHARNEY, J.G., FJØRTOFT, R. and NEUMANN, J.Von; Numerical integration of the barotropic vorticity equation. Tellus, Stockholm, 2, 1950, pp. 237-254.

3. BENWELL, G.R.R., GADD, A.J., KEERS, J.F., TIMPSON, M.S. and WHITE, P.W.; The Bushby-Timpson 10-level model on a fine mesh. Scient Pap, Met Off, London, No. 32, 1971.

4. BURRIDGE, D.M. and GADD, A.J.; The Meteorological Office operational 10-level numerical weather prediction model. Scient Pap, Met Off, London, No. 34 (to be published in 1977).

5. BURRIDGE, D.M.; A split semi-implicit reformulation of the Bushby-Timpson 10-level model. Q J R Met Soc, London, 101, 1975, pp. 777-792.

6. CORBY, G.A.; Some experiments in the objective analysis of contour charts. Q J R Met Soc, London, 87, 1961, pp. 34-42.

7. ATKINS, M.J.; The objective analysis of relative humidity. Tellus, Stockholm, 26, 1974, pp. 663-671.

8. DIXON, R.; Orthogonal polynomials as a basis for objective analysis. Scient Pap, Met Off, London, No. 30, 1969.

9. DIXON, R. and SPACKMAN, E.A.; The three-dimensional analysis of meteorological data. Scient Pap, Met Off, London, No. 31, 1970.

10. THOMPSON, P.D.; Numerical weather analysis and prediction. New York, Macmillan, 1961.

11. WEXLER, R.; Precipitation growth in stratiform clouds. Q J R Met Soc, London, 78, 1952, pp. 363-376.

12. RICKETTS, J.N.; An investigation into a relationship between upper-air relative humidity and cloud cover, Met Mag, London, 102, (May) 1973, pp. 146-156.

APPENDIX

Orthogonal polynomials - a simple example

1. Consider the set of points, $x_i = i$, where $i = 1, 2, 3$: a set of orthogonal polynomials may be derived, namely

$$\left.\begin{aligned} f_0(x) &= a_0 & \ldots \\ f_1(x) &= b_0 + b_1 x & \ldots \\ f_2(x) &= c_0 + c_1 x + c_2 x^2 & \ldots \end{aligned}\right\} \qquad \ldots . \text{(A.1)}$$

Writing $\sum_{i=1}^{3} f_m(x_i)\, f_n(x_i)$ as $\Sigma f_m f_n$, the conditions for orthogonality are

$$\Sigma f_0 f_1 = \Sigma f_0 f_2 = \Sigma f_1 f_2 = 0 \qquad \ldots . \text{(A.2)}$$

while for a non-trivial solution

$$\Sigma f_0^2, \ \Sigma f_1^2 \text{ and } \Sigma f_2^2 \neq 0 . \qquad \ldots . \text{(A.3)}$$

The orthogonality condition, as will be seen below, leads to n simultaneous equations for a polynomial with $(n+1)$ coefficients: this means that absolute values cannot be assigned to the coefficients, but that one coefficient in each polynomial must be given an arbitrary value, the remaining coefficients being determined relative to the first.

2. There is no orthogonality condition involving a_0 alone, for example, and so a_0 may be given an arbitrary value, subject to the condition that $\Sigma f_0^2 \neq 0$, i.e. $a_0 \neq 0$. Let $a_0 = f_0 = 1$. Then equation A.2 for f_0 and f_1 leads to

$$\begin{aligned} \Sigma f_0 f_1 &= 0 \\ &= \sum_{i=1}^{3} a_0 (b_0 + b_1 x_i) \\ &= a_0(b_0 + b_1) + a_0(b_0 + 2b_1) + a_0(b_0 + 3b_1) \\ &= a_0(3b_0 + 6b_1) \\ &= 3(b_0 + 2b_1) \\ b_0 &= -2b_1 . \end{aligned}$$

Let $b_0 = -2$, then $b_1 = 1$, and $f_1 = 2 - x$.

The equations for c_0, c_1 and c_2 are derived from the summed products $\Sigma f_0 f_2 = \Sigma f_1 f_2 = 0$, which give $c_0 : c_1 : c_2$ as $10 : -12 : 3$, i.e. $f_2 = 10 - 12x + 3x^2$.

The polynomials, and their values for $x = 1$, 2 and 3, are summarized below:

	x	1	2	3
f_0	1	1	1	1
f_1	$2 - x$	1	0	-1
f_2	$10 - 12x + 3x^2$	1	-2	1

3. The polynomials can also be derived by the Gram-Schmidt orthogonalization process, as described in equations 7a to 7e on page 2 of Reference 9. In the notation used above, the equations become:

$$\left.\begin{aligned} \mathbf{f}_0 &= 1 \\ \mathbf{f}_1 &= \mathbf{x} - \frac{(\mathbf{f}_0 \cdot \mathbf{x})\,\mathbf{f}_0}{(\mathbf{f}_0 \cdot \mathbf{f}_0)} \\ \mathbf{f}_2 &= \mathbf{x}^2 - \frac{(\mathbf{f}_0 \cdot \mathbf{x}^2)\,\mathbf{f}_0}{(\mathbf{f}_0 \cdot \mathbf{f}_0)} - \frac{(\mathbf{f}_1 \cdot \mathbf{x}^2)\,\mathbf{f}_1}{(\mathbf{f}_1 \cdot \mathbf{f}_1)} \end{aligned}\right\} \quad \ldots . \text{(A.4)}$$

where the bold type represents a vector in the matrix sense, a one-dimensional array of numbers.

For $i = 1, 2, 3$ we have:

$$\begin{aligned} f_0 &= 1 \\ f_1 &= x - \frac{\Sigma f_0 x_i}{\Sigma f_0^2} \cdot f_0 \\ &= x - \frac{(1\times 1) + (1\times 2) + (1\times 3)}{1+1+1} \cdot 1 \\ &= x - \frac{6}{3} \\ &= x - 2. \end{aligned}$$

(The change of sign of f_1 compared with that of paragraph 2 above is immaterial - it is only the relative magnitude of the coefficients that matters.)

$$\begin{aligned} f_2 &= x^2 - \frac{\Sigma f_0 x_i^2}{\Sigma f_0^2} \cdot f_0 - \frac{\Sigma f_1 x_i^2}{\Sigma f_1^2} \cdot f_1 \\ &= x^2 - \frac{(1\times 1) + (1\times 4) + (1\times 9)}{1+1+1} \cdot 1 - \frac{(-1\times 1) + (0\times 4) + (1\times 9)}{1+1} \cdot (x-2) \\ &= x^2 - \frac{14}{3} - \frac{8}{2}(x-2) \\ &= x^2 - 4x + \frac{10}{3}, \end{aligned}$$

equivalent to the equation given in paragraph 2.

4. Let $h(x)$ be a quantity which varies with x, and let its values at $x = 1$, 2 and 3 be 10, 6 and 5 respectively. Put $h(x) = \alpha_0 f_0(x) + \alpha_1 f_1(x) + \alpha_2 f_2(x)$. Then, substituting for the values of f_0, f_1 and f_2 from the table of paragraph 2, we have:

$$\begin{aligned} 10 &= \alpha_0 + \alpha_1 + \alpha_2 \\ 6 &= \alpha_0 \qquad - 2\alpha_2 \\ 5 &= \alpha_0 - \alpha_1 + \alpha_2 \end{aligned}$$

giving

$$\begin{aligned} \alpha_0 &= 7 \\ \alpha_1 &= 2.5 \\ \alpha_2 &= 0.5 \end{aligned}$$

whence $h(x) = 17 - 8.5x + 1.5x^2$.

Alternatively, from equation 3.10 on page 18 (see also the derivation in paragraph 5 below),

$$\alpha_0 = \frac{\Sigma b f_0}{\Sigma f_0^2} = \frac{10+6+5}{3} = 7$$

$$\alpha_1 = \frac{\Sigma b f_1}{\Sigma f_1^2} = \frac{10-5}{2} = 2.5$$

and

$$\alpha_2 = \frac{\Sigma b f_2}{\Sigma f_2^2} = \frac{10-12+5}{6} = 0.5.$$

5. <u>Least-squares derivation of α_n</u>

Let b be the true value of $b(x)$ at a given point, and $E(b)$ the 'fitted' value, i.e. considering only the first three terms,

$$E(b) = \alpha_0 f_0 + \alpha_1 f_1 + \alpha_2 f_2.$$

Then ε, the error of the estimate, is given by

$$\begin{aligned}
\varepsilon &= E(b) - b \\
&= (\alpha_0 f_0 + \alpha_1 f_1 + \alpha_2 f_2) - b \\
\varepsilon^2 &= \left[(\alpha_0 f_0 + \alpha_1 f_1 + \alpha_2 f_2) - b\right]^2 \\
&= \alpha_0^2 f_0^2 + \alpha_1^2 f_1^2 + \alpha_2^2 f_2^2 + \\
&\quad + 2(\alpha_0 \alpha_1 f_0 f_1 + \alpha_0 \alpha_2 f_0 f_2 + \alpha_1 \alpha_2 f_1 f_2) - \\
&\quad - 2b(\alpha_0 f_0 + \alpha_1 f_1 + \alpha_2 f_2) + b^2.
\end{aligned}$$

Summing over all values:

$$\begin{aligned}
\Sigma\varepsilon^2 &= \Sigma\alpha_0^2 f_0^2 + \Sigma\alpha_1^2 f_1^2 + \Sigma\alpha_2^2 f_2^2 - \\
&\quad - 2\Sigma b(\alpha_0 f_0 + \alpha_1 f_1 + \alpha_2 f_2) + \Sigma b^2,
\end{aligned}$$

since $\Sigma f_0 f_1 = \Sigma f_0 f_2 = \Sigma f_1 f_2 = 0$ by the orthogonality condition.

The least-squares condition is that $\Sigma\varepsilon^2$ is a minimum, i.e.

$$\frac{\partial(\Sigma\varepsilon^2)}{\partial\alpha_0} = -2\,\Sigma b f_0 + 2\alpha_0 \Sigma f_0^2 = 0$$

$$\frac{\partial(\Sigma\varepsilon^2)}{\partial\alpha_1} = -2\,\Sigma b f_1 + 2\alpha_1 \Sigma f_1^2 = 0$$

and

$$\frac{\partial(\Sigma\varepsilon^2)}{\partial\alpha_2} = -2\,\Sigma b f_2 + 2\alpha_2 \Sigma f_2^2 = 0,$$

giving

$$\alpha_0 = \frac{\Sigma b f_0}{\Sigma f_0^2}$$

with similar expressions for α_1 and α_2. The expressions for the αs are analogous to those found in simple linear regression, for, if $y = bx + c$, the least-squares estimate of b is $\Sigma xy / \Sigma x^2$.

Differentiation a second time shows that $\dfrac{\partial^2(\Sigma\varepsilon^2)}{\partial\alpha_n^2} = 2\Sigma f_n^2 > 0$, and the above conditions are therefore for $\Sigma\varepsilon^2$ to be a minimum.