INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR/534-2

ICTP/WMO WORKSHOP ON EXTRA-TROPICAL AND TROPICAL
LIMITED AREA MODELLING
22 October - 3 November1990

"Meteorological data analysis"

by

P. Lönnberg (Chapters 1-5)
A. Hollingsworth (Chapter 6)

Presented by:
Prof. S. TIBALDI
Institute of Physics
Univ. of Bologna
Bologna, Italy

Please note: These are preliminary notes intended for internal distribution only.

European Centre
for Medium Range Weather Forecasts

LECTURE NOTE

METEOROLOGICAL TRAINING COURSE

Lecture Note No. 2.2

METEOROLOGICAL DATA ANALYSIS

by

P. Lönnberg (Chapters 1-5)
A. Hollingsworth (Chapter 6)

Acknowledgement

In preparing these notes on meteorological data analysis the authors have made extensive use of an earlier set of notes by A. Lorenc on the same subject.

February 1984

CONTENTS

	PAGE NO.
1. INTRODUCTION	1
1.1 Summary	1
1.2 Aims	1
1.3 Data available	2
1.4 Geostrophic adjustment as a function of scale	4
2. LEAST SQUARES BEST FIT TECHNIQUES	11
2.1 Method of least squares	11
2.2 Mathematical functions	12
2.2.1 Local fit by polynomials	12
2.2.2 Global fit to periodic functions	13
2.3 Meteorological functions	13
2.3.1 Empirical orthogonal functions (EOFs)	13
2.3.2 Normal mode expansion	14
3. EMPIRICAL CORRECTION METHODS	17
4. TECHNIQUES TO EXTRACT AND COMBINE ANALYSIS FIELDS CONSISTENTLY	21
4.1 Use of a balance equation	21
4.2 Variational methods	22
5. STATISTICAL METHODS	23
5.1 Concepts and notation	23
5.2 Optimal interpolation (OI)	24
5.3 Simple example of OI	28
5.4 Prediction error statistics	29
5.4.1 Horizontal forecast error correlations	30
5.4.2 Vertical forecast error correlations	34
6. SPECTRAL RESPONSE OF AN OPTIMAL INTERPOLATION ANALYSIS	37
6.1 Spectral response of the OI analysis	37
6.2 Some simple three point analysis problems	41
6.3 Numerical examples - OI as a smoothing/desmoothing operator	73
6.3.1 Univariate analysis with a Gaussian correlation - Numerical example.	73
6.3.2 Effect of observation spacing or width of structure function	74
6.3.3 A finite difference interpretation of the OI results	75
6.3.4 Numerical example - univariate wind analysis	76
6.4 Discussion	81
REFERENCES	85

1. INTRODUCTION

1.1 Summary

This course is intended to give insight to methods employed in meteorological data analysis. The first section outlines the aims of the analysis, the data available, and introduces some notations and terminology. The first numerical methods of objective data analysis were based on the least squares best fit technique (Section 2). The computational cost and meteorological instability of the least squares methods were overcome by empirical (correction) methods (Section 3). Section 4 describes how diagnostic relationships between meteorological variables can be used to derive consistent analyses. The statistical interpolation method which is currently in wide use is described in detail in Section 5. Some simple examples are worked out in detail in Section 6.

1.2 Aims

A numerical description of the atmosphere is needed for :

1. Initializing a numerical forecast. In this case one can say that the 'best' analysis is the one which gives the best forecast. Hence the analysis scheme may need tuning for the particular forecast model to be used.

2. Diagnostic studies of the atmosphere, plotting of charts,forecast verification. The 'best' analysis in this case may be different from that for 1. Systematic biases need to be avoided for diagnostic and climatological studies. An analysis which uses a forecast as first-guess cannot easily be used to verify that forecast at short forecast range.

3. Observation checking. For all practical analysis schemes this is a very important problem. Some checking of observations is possible internally within one observation, or by comparison with climatology or a forecast, but the best final check is against an analysis made

without the datum being checked. Thus provision must be made either within the analysis process or after a preliminary scan to check the data used for the analysis.

1.3 Data available

1. Observations. These have been described in earlier lectures.

2. "First-guess". Most analysis methods use a preliminary estimate of the field being analyzed (returned more or less unchanged as the analysis if there are no observations). Usually this is done by analysing deviations from the first-guess rather than the total values. The fields available for use as a first-guess are

a) climatology, i.e. the average of previous analyses for the same season

b) persistence, i.e. the previous analysis

c) forecast, i.e. a numerical prediction from the previous analysis.

These fields, if they are available, have to be combined to give one first-guess field. The optimal combination is scale-dependent, and the analysis method needs to be tuned for the type of first-guess used.

3. Knowledge of the likely structure and scale of atmospheric motions. This information is often incorporated implicitly into analysis schemes without being clearly stated or quantified. Other knowledge is expressed as relationships which the atmosphere (approximately) obeys, and which are used as either weak or strong constraints on the analysis. The following statements have all been used in analysis schemes.

a) Increments to atmospheric fields are smooth and continuous. All schemes we know of assume this, despite the existence of fronts.

b) The value being analysed (usually the first-guess error) is most likely to have a certain scale. Thus a single isolated observation will cause the analysis to be a feature of this most likely scale, and dense inaccurate observations will be averaged over this scale in an ideal analysis scheme.

c) The atmosphere is in hydrostatic balance.

d) The atmosphere is in geostrophic or gradient wind balance.

e) The horizontal wind is non-divergent.

f) The atmosphere is in a state which satisfies the balance equation.

g) The atmosphere is convectively stable.

h) The atmosphere is not super-saturated.

Note however that there are still some properties of the atmosphere which it is possible to explain, and use in a subjective analysis, but which are difficult to use in a numerical scheme:

i) Mid-latitude systems often have the characteristic shape of a warm sector depression.

j) Certain regions are preferred for the development of new, initially small scale, depressions.

k) Developing systems usually have a vertical phase change (tilt).

1.4 Geostrophic adjustment

The atmospheric velocity field is generally close to a state of geostrophic balance and consequently, the total energy associated with non-geostrophic motions must be relatively small. In the real atmosphere, high frequency non-geostrophic phenomena may be of local importance but on a global scale they must be relatively insignificant compared to the sources that maintain the quasi-balanced state. A global build-up of energy of locally excited high-frequency waves is inhibited by mechanisms such as frictional dissipation, vertical energy flux, and nonlinear interaction.

A numerical forecast model with only a limited number of degrees of freedom cannot describe the geostrophic adjustment outlined above. Therefore, the high-frequency waves of the initial state cannot be dissipated locally by small-scale phenomena. The initial conditions can by a proper initialisation procedure be determined so that there is no noise in the forecast. Although a balanced state can be found for the forecast model it is essential that the analysis depicts the Rossby modes as accurately as possible to avoid rejection of information by the initialisation.

In the following we will investigate how a simple model reacts to an imbalance in the initial state and how its response depends on the scale of motion. To damp the non-geostrophic modes we dissipate the divergent wind and consequently the stationary state of the model is in geostrophic balance. The linearized adjustment theory applied to a two-dimensional situation (Temperton, 1973) is described below.

Let us investigate the simplified equations for a homogeneous incompressible fluid. Let us suppose that the fields u, v, and h can be represented as a sum of two-dimensional Fourier components, e.g.,

$$u(x,y,t) = \sum_k \sum_\ell u_{k\ell}(t) \exp\,[i(kx + \ell y)] \qquad (1.4.1)$$

Since we are studying a linearized system we can consider a particular pair of wavenumbers. For convenience we will also drop the subscripts from the Fourier coefficients. The transformed system thus becomes

$$\frac{\partial u}{\partial t} = fv - ikgh \qquad (1.4.2)$$

$$\frac{\partial v}{\partial t} = -fu - i\ell gh \qquad (1.4.3)$$

$$\frac{\partial h}{\partial t} = -iD_O(ku + \ell v) \qquad (1.4.4)$$

where D_O is the mean geopotential height of the free surface.

We will next separate the wind field into a nondivergent part (ψ) and a nonrotational part (χ). This gives

$$u = -i(\ell\psi - k\chi)$$

$$v = i(k\psi + \ell\chi) \qquad (1.4.5)$$

Inserting Equation (1.4.5) into Equations (1.4.2)-(1.4.4) yields

$$\frac{\partial \psi}{\partial t} = -f\chi \qquad (1.4.6)$$

$$\frac{\partial \chi}{\partial t} = f\psi - gh \qquad (1.4.7)$$

$$\frac{\partial h}{\partial t} = D_0(k^2 + \ell^2)\chi \qquad (1.4.8)$$

Equation (1.4.6) corresponds to the vorticity equation, and Equation (1.4.7) corresponds to the divergence equation, and (1.4.8) to the continuity equation.

As can be seen from the solutions to this system, it is not possible to describe geostrophic adjustment since the solution for the divergence can be expressed by the simple wave equation, easily obtained from Equations (1.4.6)-(1.4.8)

$$\frac{\partial^2 \chi}{\partial t^2} + \omega^2 \chi = 0 \qquad (1.4.9)$$

where

$$\omega_{1,\,2} = \pm \sqrt{f^2 + gD_0(k^2 + \ell^2)} \qquad (1.4.10)$$

and hence the general solution to (1.4.9) is

$$\chi = Ae^{i\omega_1 t} + Be^{i\omega_2 t} \qquad (1.4.11)$$

Here the gravity waves are affected by the rotation of the earth and therefore are usually called inertia-gravity waves. Inertia-gravity waves are dispersive, that is, the phase speed expressed by ω is a function of wavelength. This is not the case with pure gravity waves which move with the same speed independent of wavelength. The dispersive nature of inertia-gravity waves provides another mechanism for locally damping waves in the atmosphere. However, in order to describe this process we have to solve a nonlinear problem.

In order to simulate the effect of dispersion and interference of gravity waves, we will introduce a viscous damping term acting on the divergent wind only. This term will have the form $\kappa\nabla^2\chi$. We thus obtain the new system

$$\frac{\partial \psi}{\partial t} = -f\chi \qquad (1.4.12)$$

$$\frac{\partial \chi}{\partial t} = f\psi - gh - \kappa(k^2 + \ell^2)\chi \qquad (1.4.13)$$

$$\frac{\partial h}{\partial t} = D_0(\ell^2 + k^2)\chi \qquad (1.4.14)$$

We will now consider a stationary state of the modified Equations (1.4.12)-(1.4.14). Since then $\frac{\partial \psi}{\partial t} = \frac{\partial \chi}{\partial t} = \frac{\partial h}{\partial t} = 0$, we must have the relations

$$\chi_s = 0$$

$$h_s = \frac{f}{g}\psi_s \qquad (1.4.15)$$

where subscript s indicates the stationary case. Equation (1.4.15) shows that in the stationary case the geopotential and the wind fields are in geostrophic balance. It is found that an invariant quantity, analogous to the potential vorticity, exists in both the original and modified form of the equations

$$\Omega = D_0(k^2 + \ell^2)\psi + fh, \quad \frac{\partial \Omega}{\partial t} = 0 \qquad (1.4.16)$$

If the system (Equations 1.4.12-1.4.14) is now solved as an initial-value problem with the initial fields not in geostrophic balance, the initial and stationary values of Ω can be related by $\Omega_s = \Omega_i$ (subscript i indicates the initial value).

$$D_0(k^2 + \ell^2)\psi_s + fh_s = D_0(k^2 + \ell^2)\psi_i + fh_i \qquad (1.4.17)$$

Inserting the stationary values given in Equation (1.4.15) yields

$$[D_0(k^2 + \ell^2) + f^2/g]\,\psi_s = D_0(k^2 + \ell^2)\psi_i + fh_i$$

and finally

$$\psi_s = \alpha\psi_i + (1-\alpha)\,gh_i/f \qquad (1.4.18)$$

where gh_i/f is an expression for the geostrophic balance with the initial geopotential field and

$$\alpha = \frac{gD_0(k^2 + \ell^2)}{gD_0(k^2 + \ell^2) + f^2} \qquad (1.4.19)$$

α can vary from 0 when the characteristic dimensions of the perturbations are very large (k,ℓ very small) and $f > 0$ to 1 for a small scale perturbation (k,ℓ very large) or when $f = 0$.

If $\alpha \to 1$ then $\psi_s \approx \psi_i$ and $h_s = (f/g)\psi_i$, i.e., the mass field adjusts to the initial wind field, and if $\alpha \to 0$ then $\psi_s = gh_i/f$ and the wind field adjusts to the initial mass field.

The physical significance of Eqn.(1.4.18) is made clearer if we transform it back to Cartesian space. Noting that $k = 2\pi/\lambda_x$ and $\ell = 2\pi/\lambda_y$ for the wavelengths λ_x and λ_y in the x- and y- directions respectively we can write

$$\alpha = \lambda_c^2/(\lambda_c^2 + \lambda_{xy}^2)$$

where $\lambda_c = 2\pi R_c$ is called the critical wavelength. R_c is the Rossby radius of deformation, $R_c = (gD_o)^{1/2}/f$ and

$$\lambda_{xy} = \left(\frac{1}{\lambda_x^2} + \frac{1}{\lambda_y^2}\right)^{-1/2} = \frac{\lambda_x \lambda_y}{(\lambda_x^2 + \lambda_y^2)^{1/2}} \qquad (1.4.20)$$

is called the effective length scale. Now the conditions become the following: if $\lambda_{xy} < \lambda_c$ (small-scale waves), the mass field adjusts to the wind field while if $\lambda_{xy} > \lambda_c$ (large-scale waves), the wind field adjusts to the mass field.

It is important to observe that for an effective scale, λ_{xy}, to be greater than the critical wavelength, λ_c, we must have both $\lambda_x > \lambda_c$ and $\lambda_y > \lambda_c$. In other words, no matter how great the wavelength is in one direction, if the wavelength in the other direction is shorter than the critical wavelength, the system will behave as if it were small-scale.

The implication of this analysis for the process of dynamical initialization is that forcing the wind field to adjust to the mass field is "unnatural" for the "small scale" components involved. Consider the particular case when $\alpha \approx 1$, $\psi_s \approx \psi_1$, and $\chi_s \approx 0$. If in this special case the initial perturbations

were applied to the mass field only ($\psi_i = \chi_i = 0$), the fluid would still be completely at rest in the final state, ψ_s. That is, there has been no assimilation of the mass field perturbation whatsoever and this perturbation has been dissipated entirely in the form of gravity waves. Naturally this fact has great significance for the problem of data assimilation.

2. LEAST SQUARES BEST FIT TECHNIQUES

A widely used procedure in geophysics is the method of least squares due to Gauss. The basic principle is to fit a set of functions to observed data. This technique has been applied to analysis problems in a variety of versions. We group these, somewhat artifically, according to the type of functions used to fit the observations. Examples of local and global applications with extensions to several variables are presented.

2.1 Method of least squares

In this method a linear combination of M functions F_k is fitted to a (generally) irregular set of N observations A_j^o at positions x_j. The representativeness of each observation is given by a weight w_i which is determined from the observation error and data density. The coefficients C_k of the analysis field

$$A^i(x) = \sum_{k=1}^{M} C_k F_k(x) \qquad (2.1.1)$$

are solved by minimizing the sum of the squared differences between analysis and observations

$$S = \sum_{j=1}^{N} w_j (A_j^o - A_j^i)^2 \qquad (2.1.2)$$

with respect to C_k. This leads to a system of M linear equations

$$\frac{\partial S}{\partial C_\ell} = 0 \text{ for } \ell = 1, \ldots, M \qquad (2.1.3)$$

Substitution of (2.1.2) into (2.1.3) gives

$$\sum_{j=1}^{N} w_j \{F_\ell(x_j) \sum_{k=1}^{M} C_k F_k(x_j)\} = \sum_{j=1}^{N} w_j A_j^o F_\ell(x_j) \qquad (2.1.4)$$

for $\ell=1, \ldots, M$.

If the functions are orthogonal with respect to the weighting and data distribution, (2.1.4) reduces to

$$C_\ell = \sum_{j=1}^{N} w_j A_j^{\circ} F_\ell(x_j) / \sum_{j=1}^{N} w_j F_\ell^2(x_j) \qquad (2.1.5)$$

for $\ell = 1,\ldots,M$

2.2 Mathematical functions

Under this heading we collect least squares analysis methods that are based on functions we consider non-meteorological. Functions which describe the atmospheric structure or which are solutions to equations of the atmospheric state are discussed in Section 2.3

2.2.1 Local fit by polynomials

Gilchrist and Cressman (1954) constructed a local least squares analysis scheme in which the geopotential is expressed as a polynomial of 2nd degree.

$$z(x,y) = \sum_{i,j} c_{ij} x^i y^j \qquad \begin{matrix} i,j \geq 0 \\ \text{and} \\ i+j \leq 2 \end{matrix} \qquad (2.2.1)$$

The gridpoint to be analysed is the origin of the coordinate system and consequently the coefficient c_{oo} is the analysed geopotential.

Wind observations can be included geostrophically in this method.
The six coefficients c_j are solved by minimization of

$$S = \frac{1}{\delta_z} \sum_{r=1}^{R} (z_r^{obs} - z_r^{an})^2 + \frac{1}{\delta_w} \sum_{s=1}^{S} \left\{ \left(u_s^{obs} + \frac{g}{f}\frac{\partial z_s^{an}}{\partial y}\right)^2 + \left(v_s^{obs} - \frac{g}{f}\frac{\partial z_s^{an}}{\partial x}\right)^2 \right\} \qquad (2.2.2)$$

in which δ_z and δ_w are the estimated observation errors for height and wind components respectively.

In a polynomial type analysis of degree n, the geostrophic wind variations are of degree n-1. Consequently, for n=2 the wind variations in the data domain are assumed to be linear. Higher order versions of this technique easily overfit the observations and produce large oscillations in data sparse regions. The number of coefficients to be determined is (n+1) (n+2)/2.

2.2.2 Global fit to periodic functions

Periodic functions are a natural choice for the representation of meteorological fields on a cyclic domain. The influence on the analysis of each observation extends over the whole domain which might be the entire globe. The noise in the observations is difficult to separate from the meteorological signal but it can be removed by a proper spectral truncation of the functions. Moreover, a non-uniform data distribution might excite modes which are not present in the observations.

In some applications orthogonality of the basis functions is exploited to reduce the least squares problem to form (2.1.5).

2.3 Meteorological functions

The least squares techniques we discussed in the previous section did not exploit any knowledge of the meteorological behaviour of the atmosphere. Functions describing the long-term statistical variability of the atmosphere as well as solutions to the governing equations have been used as basis functions.

2.3.1 Empirical orthogonal functions (EOF's)

The statistical structure of the atmosphere is conveniently expressed by a covariance matrix of observed or analysed departures from climatology at a number of spatial positions. The covariance matrix can be expanded into its eigenvalues and eigenvectors (EOF's) and the relative magnitude of a particular eigenvalue gives the contribution of that mode to the total

variance. The eigenvectors related to the gravest modes are chosen as basis functions for the least squares method. EOF's have been mainly used to define the vertical structure of the atmosphere. Several applications of EOF's can be found in the Proceedings of the ECMWF Workhop on the use of empirical orthogonal functions in meteorology (1977).

2.3.2 Normal mode expansion

All the analysis methods described so far suffer from lack of global balance between height and wind fields. Solutions to various approximations of the governing equations provide sets of functions with some balance properties. Spherical harmonics (solutions of a linearized non-divergent vorticity equation) and Hough functions (solutions to the "shallow water" equations) are the most commonly used approximations to the atmospheric flow in global data fitting.

In the following we will only consider the Hough function expansion. These functions are solutions to the linearized primitive equations on a sphere for a basic state at rest and with temperature as a function of height only. Furthermore, the flow is assumed to be inviscid and incompressible and the terrain flat. We assume solutions of the form (Kasahara, 1976)

$$Y = \begin{bmatrix} u \\ v \\ h \end{bmatrix} = \sum_{n=s}^{\infty} \begin{bmatrix} A_n^s \\ iB_n^s \\ C_n^s \end{bmatrix} P_n^s(\mu) \exp(is\lambda - i\sigma t) \qquad (2.3.1)$$

in which s is the zonal wave number,

$P_n^s(\mu)$ is the associated Legendre polynomial of the first kind of degree s and order n, $\mu = \sin\phi$ and σ is the frequency.

In practice, the infinite series in (2.3.1) is limited to N terms. Substitution of the finite series into the "shallow-water" equations leads to an eigenvalue problem of 3N eigenvalues and eigenvectors. The frequency of a mode is determined by its eigenvalue and the horizontal structure, i.e.

efficients A_n^s, B_n^s and C_n^s, by its eigenvector. The meridional part of a mode ℓ and zonal wavenumber s is

$$\Theta_\ell^s(\phi) = \sum_{n=s}^{s+N} \begin{bmatrix} A_n^s \\ iB_n^s \\ C_h^s \end{bmatrix} P_n^s(\sin\phi) \qquad (2.3.2)$$

θ_ℓ^s is called the Hough vector function.

The 3N modes include N low-frequency Rossby modes and 2N gravity modes.

Flattery (1971) expanded meteorological observations in terms of Hough harmonics $\theta_\ell^s \exp(\pm is\lambda)$. The Hough vector functions are determined for an equivalent depth of 10 km. The vertical structure of atmosphere is defined by 7 EOF's. The u,v and z analysis can then be written as follows

$$\begin{bmatrix} u \\ v \\ h \end{bmatrix}(\lambda,\phi,p) = \sum_{\ell=o}^{24} \sum_{s=1}^{24} \sum_{n=1}^{7} \{a_{\ell sn} \cos(s\lambda) + b_{\ell sn} \sin(s\lambda)\} \Theta_\ell^s(\phi).E_n(p) \qquad (2.3.3)$$

The expansion coefficients $a_{\ell sn}$ and $b_{\ell sn}$ are solved by minimising

$$\epsilon I_z + I_u + I_v \qquad (2.3.4)$$

where

$$I_f = \int_V \{f^{obs} - f^{an}\} \, dV \qquad (2.3.5)$$

The weighting ϵ depends on which variable (height or wind) is being analysed. A uniform data distribution is achieved by forming average observations in 3° x 6° boxes. In empty boxes a forecasted value is used. Consequently, data voids influence the analysis as strongly as data dense areas, which may lead to serious aliasing problems. Alternatively, each observation or observation box may be assigned a weight according to the accuracy of its meteorological value. In this case, the orthogonality of the basis functions cannot be used and a system of linear equations of form (2.1.4) must be

solved. In applications with variable observation weighting, the number of modes is severely limited due to computational considerations.

In least squares techniques, the measurement noise is filtered out by spectral truncation of the basis functions. Non-uniform data distribution and inaccurate observations makes data expansion in normal modes or other global functions feasible only for low wavenumbers.

3. EMPIRICAL CORRECTION METHODS

The local least squares methods suffer from an inherent problem related to the number of approximating functions. An economical and straightforward method was proposed by Bergthorsson and Döös (1955) which overcame the overshooting problems of polynomial fitting. The empirical correction method performs well in two dimensions for homogeneous observations. The principle of the method is to modify a guess field locally by observations as follows:

A. Construct the first-guess field from a forecast and climatology

$$A_i^p = \frac{\alpha_i^{fc} A_i^{fc} + \alpha_i^{cl} A_i^{cl}}{\alpha_i^{fc} + \alpha_i^{cl}} \qquad (3.1)$$

The weighting factors α are defined by the accuracy of the contributing fields

$$\alpha_i^{fc} = (E^{fc})^{-1} \text{ and } \alpha_i^{cl} = (E^{cl})^{-1} \qquad (3.2)$$

in which E is the std of the error of the forecast or climatology. The symbol A represents as before the meteorological value.

B. Calculate a weight for each observation influencing the analysis at gridpoint k.

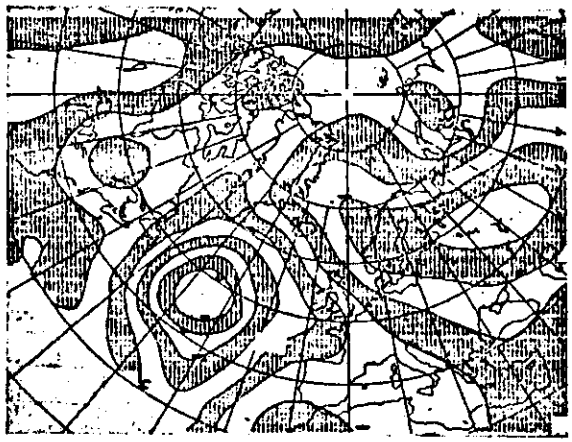
The empirical part of the method is the specification of the influence (weight) function. The weights depend on the distance to the analysis point, data density and observation errors. Usually, the weights are normalised by the sum of the weights and the accuracy of the first-guess.

C. Calculate the correction to the first-guess

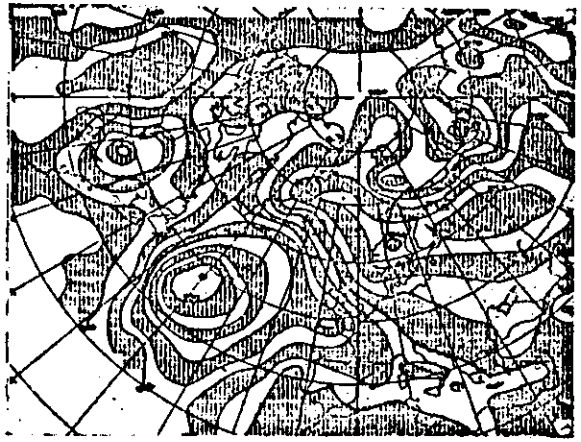
$$A_k^i - A_k^p = \sum_i w_{ki} (A_i^o - A_i^p) \qquad (3.3)$$

D. Steps B and C may be repeated with the analysis as the guess field and a different (sharper) influence function ("successive correction method"). This technique was proposed by Cressman (1959). An example of a successive correction analysis is shown in Fig. 3.1

Wind information accompanied by height data provide additional estimates of the height at the gridpoint through extrapolation using the geostrophic or gradient wind relation. The curvature of the flow is taken from the guess field.



First scan in an analysis of the surface pressure. The radius of influence (N) of the stations in this scan is 6 grid units = 1800 km.



Fifth scan. N = 1 grid unit = 300 km.

Fig. 3.1 Scans 1 and 5: This figure shows an example of the analysis of surface pressure using a method of successive scans. Only scans 1 and 5 are shown. In each successive scan the area influencing each station was decreased. Winds were introduced into the analysis in the last three scans only. Note the change in the refinement in the analysis between scans 1 and 5 (From Döös, 1969).

4. TECHNIQUES TO EXTRACT AND COMBINE ANALYSIS FIELDS CONSISTENTLY

Diagnostic relations between meteorological parameters provide means to extract new fields from analysed variables. Section 4.1 briefly describes the use of the balance equation to retrieve wind or height from the other field. In section 4.2 we discuss variational techniques to combine separately analysed fields under given constraints.

4.1 Use of a balance equation

In extratropical regions where the quality of the height analysis is generally high, the wind field can be determined from the geopotential by some form of the balance equation:

$$\nabla^2\phi = f\nabla^2\psi \qquad \text{quasi-geostrophic relation}$$

$$\nabla^2\phi = \nabla.(f\nabla\psi) \qquad \text{linear balance equation}$$

$$\nabla^2\phi = 2J\left(\frac{\partial\psi}{\partial x}, \frac{\partial\psi}{\partial y}\right) + \nabla.(f\nabla\psi) \qquad \text{balance equation}$$

$$\nabla^2\phi = \eta\zeta + \kappa.\nabla\zeta \times \underline{v} - \nabla^2(v^2/2) \qquad \text{vorticity form of balance equation}$$

in which

$$\zeta = \kappa.\nabla \times \underline{v} = \nabla^2\psi \quad \text{and} \quad \eta = f + \zeta \ .$$

Local non-ellipticity of the height field occasionally creates convergence problems.

In the tropics the climatological variability of the height field is of the same order as the observation error. The geopotential can, however, be derived with reasonable accuracy by the balance equation from the streamfunction field. This in turn can be solved through the Poisson equation $\nabla^2\psi = \zeta$ from the vorticity of the analysed wind.

4.2 Variational methods

Variational techniques can be applied to combine independently analysed fields when a constraint between them is required. Already very simple applications of variational calculus become mathematically cumbersome and consequently we will only present the principles of the method. The idea is to find the minimum changes to be made to the separately analysed fields in order to satisfy the imposed constraint either fully (strong constraint) or approximately (weak constraint).

Suppose that we have analyses of z, u, v, T and q (denoted by a superscript a). The problems are then the following:

A. Strong constraint (Sasaki, 1958)

Minimize

$$I = \int_V \left[\alpha_z (\Delta z)^2 + \alpha_u (\Delta u)^2 + \alpha_v (\Delta v)^2 + \alpha_T (\Delta T)^2 + \alpha_q (\Delta q)^2\right] dV \qquad (4.2.1)$$

in which Δf are the changes to f^a and the α's are measures of the accuracy of the respective analyses. The constraint

$$F(z^a + \Delta z,\ u^a + \Delta u, \ldots) = 0 \qquad (4.2.2)$$

must be satisfied everywhere in the domain V.

B. Weak constraint (Sasaki, 1970)

Minimize

$$I = \int_V \left[\alpha_z (\Delta z)^2 + \alpha_u (\Delta u)^2 + \ldots + \alpha_F F^2\right] dV \qquad (4.2.3)$$

The constraint F is allowed to be nonzero and its effectiveness is tuned by α_F.

Finding the minimum of the functional I is very time consuming even in the case where the α's are constant in the domain.

5. STATISTICAL METHODS

5.1 Concepts and notations

a) Expected value. This is the average value of a large number of realizations, all with the same constraints. This is denoted by triangular brackets $\langle\ \rangle$.

b) True value. This is the actual value after scales which we do not wish to analyze have been removed. We denote this by a super-fix ^{t}. Observed and predicted (first-guess) values are denoted by super-fixes ^{o} and ^{p} respectively. We shall assume that observations and first-guess are unbiased, i.e. that $\langle A^o \rangle = \langle A^p \rangle = A^t$.

c) Error. The error in a particular case is the deviation from the truth. We denote this by a lower case a:- $A^p - A^t = a^p$, $A^o - A^t = a^o$.
Note that since the truth is defined as containing only those scales which we wish to analyze, then an accurate observation of a small scale, e.g. a gust of wind, will be included in the error. The expected value of the error variance is denoted by E^2 :-

$$E^{p2} = \langle a^{p2} \rangle$$

$$E^{o2} = \langle a^{o2} \rangle$$

d) Covariances. For two points i and j the prediction error covariance is $\langle a_i^p a_j^p \rangle$, the observation error covariance is $\langle a_i^o a_j^o \rangle$ and the observation-prediction covariance is $\langle a_i^o a_j^p \rangle$. The prediction error covariance function defines the scales of the error we wish to analyze, since it is the Fourier transform of the error power spectrum.

e) Correlations. The error correlation μ is defined by the relationship

$$\langle a_i a_j \rangle = E_i \mu_{ij} E_j .$$

f) Homogeneity. A statistical property of a meteorological field that depends on two position vectors is homogeneous if it is independent of translation of the two positions.

g) Isotropy. A statistical property is isotropic if it is independent of rotation.

Homogeneity and isotropy together imply that the statistical property depends on distance only. Since the expected errors (E) of meteorological fields often vary spatially (e.g. as a function of latitude) it is usually a better approximation to assume that it is the correlation μ which is homogeneous and isotropic. Typical empirical correlations are shown in Fig.5.1.

5.2 Optimal interpolation

This technique is usually credited to Gandin (1963). It is optimal in the sense that the expected interpolation error variance is minimized if the first-guess error and observation error covariances are accurately known. Since this is rarely the case we prefer the name "statistical interpolation". Below we derive the equations in a dimensionless form, normalized by the expected rms first-guess errors E^p .

It is assumed that the normalized analyzed deviation from the first-guess can be expressed as a linear combination of normalized observation deviations:-

$$\frac{A_k^i - A_k^p}{E_k^p} = \sum_{i=1}^{N} W_{ki} \frac{A_i^o - A_i^p}{E_i^p} \qquad (5.2.1)$$

where subscript k denotes the point and variable being analyzed and subscripts i = 1, N range over all points and variables of observation.

W_{ki} are the weights to be determined.

Writing

$$\begin{aligned} \alpha_i^o &= (A_i^o - A_i^t)/E_i^o \\ \alpha_i^p &= (A_i^p - A_i^t)/E_i^p \\ \alpha_k^i &= (A_k^i - A_k^t)/E_k^i \end{aligned} \qquad (5.2.2)$$

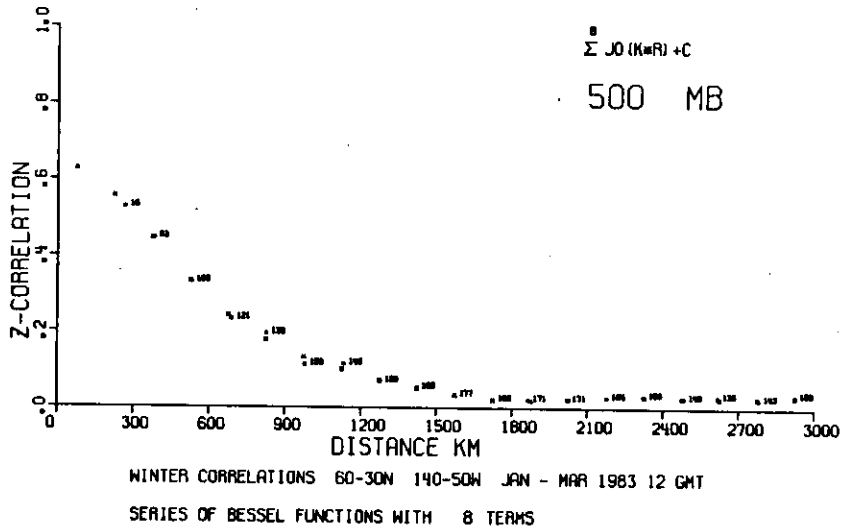


Fig. 5.1 Empirical 500mb height forecast error correlations averaged over 150km intervals (small squares). The number to the right of a square is the number of station pairs with a separation distance within a specific interval. A correlation model based on a series of Bessel functions is fitted to the interval average correlations and resulting model is indicated by a cross.

$$\varepsilon_i^o = E_i^o / E_i^p$$
$$\varepsilon_k^i = E_k^i / E_k^p$$

The interpolation equation becomes

$$\alpha_k^i \varepsilon_k^i = \alpha_k^p + \sum_{i=1}^{N} W_{ki} \, (\alpha_i^o \varepsilon_i^o - \alpha_i^p) \qquad (5.2.3)$$

Squaring this, and taking the ensemble average (denoted by < >) gives

$$\varepsilon_k^{i2} = 1 + 2 \sum_{i=1}^{N} W_{ki} \, (\langle \alpha_k^p \alpha_i^o \rangle \varepsilon_i^o - \langle \alpha_k^p \alpha_i^p \rangle)$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ki} (\langle \alpha_i^p \alpha_j^p \rangle + \varepsilon_i^o \langle \alpha_i^o \alpha_j^o \rangle \varepsilon_j^o - \varepsilon_i^o \langle \alpha_i^o \alpha_j^p \rangle \qquad (5.2.4)$$

$$- \langle \alpha_i^p \alpha_j^o \rangle \varepsilon_j^o) \, W_{kj}$$

To simplify subsequent algebra we assume at this point that the correlations of prediction and observation errors are zero, i.e.

$$\langle \alpha_i^o \alpha_j^p \rangle = \langle \alpha_i^p \alpha_j^o \rangle = 0 \qquad (5.2.5)$$

This assumption is reasonable for the types of observation currently available; if necessary it could be relaxed. We also introduce a vector and matrix notation:- $\underline{W}_k$ is the column vector of weights W_{ki}, $\underline{\underline{P}}$ is the prediction error correlation matrix

$$(\langle \alpha_i^p \alpha_j^p \rangle),$$

$\underline{\underline{O}}$ is the scaled observation error correlation matrix

$$(\varepsilon_i^o \langle \alpha_i^o \alpha_j^o \rangle \varepsilon_j^o) \text{ , and } \underline{\underline{M}} = \underline{\underline{P}} + \underline{\underline{O}}.$$

$\underline{B}$ is the vector of normalized increments

$$\frac{A_i^o - A_i^p}{E_i^p}$$

This gives

$$\varepsilon_k^{i2} = 1 - 2 \, \underline{W}_k^T \underline{P}_k + \underline{W}_k^T \underline{\underline{M}} \, \underline{W}_k \qquad (5.2.6)$$

The "optimal" weight vector is that which minimizes the estimated normalized interpolation error variances ε_k^{i2}. The usual minimisation procedure gives

$$\underline{W}_k = \underline{\underline{M}}^{-1} \underline{P}_k \qquad (5.2.7)$$

The minimum value of ε_k^{i2} corresponding to these weights is given by

$$\varepsilon_k^{i2} = 1 - \underline{W}_k^T \underline{P}_k \qquad (5.2.8)$$

The optimal interpolated increment B_k^i is given by

$$B_k^i = \underline{W}_k^T \underline{B}$$
$$= \underline{P}_k^T \underline{\underline{M}}^{-1} \underline{B} \qquad (5.2.9)$$

Note that $\underline{\underline{M}}^{-1} \underline{B}$ is independent of the analysis point k. Let us write $\underline{C} = \underline{\underline{M}}^{-1} \underline{B}$. Then

$$B_k^i = \underline{C}^T \underline{P}_k \qquad (5.2.10)$$

Thus the optimal analyzed field defined by different analysis points k is a linear combination of the prediction error correlation functions for each observation point.

Since statistical interpolation can provide an estimate of the error in the interpolated value, it is easy to devise objective observation checking schemes by comparing each datum in turn with a value interpolated from surrounding data.

The analysis and in particular the estimated analysis error depend strongly on the optimality of the forecast and observation error statistics as discussed by Seaman (1977) and Franke and Gordon (1983).

5.3 Simple example of OI

Before proceeding to more elaborate examples of OI (Section 6), we apply the equations of Sect. 5.2 to the case of only one influencing observation. The matrices and vectors reduce to very simple forms as follows

$$\underline{M} = \underline{P} + \underline{O} = \left[\mu_{11} + \varepsilon_1^{02}\right] = \left[1 + \varepsilon_1^{02}\right]$$

and $$\underline{P}_k = \left[\mu_{k1}\right]$$

where $\mu_{ij} = \langle \alpha_i^P \alpha_j^P \rangle$.

The inverse of $\underline{M}$ is $1/(1 + \varepsilon_1^{02})$ and the weight given to the observation is $\mu_{ki}/(1 + \varepsilon_1^{02})$.

At the observation point the increment to the first-guess is

$$\frac{b_1}{1 + \varepsilon_1^{02}} \qquad (5.3.1)$$

in which b_1 is the observed normalised departure from the first-guess. For simplicity it is assumed that the observed and analysed variables are the same, although this discussion applies to any variable combination for which μ_{k1} is defined.

The difference between observation and analysis is

$$(1 - w_{k1})b_1 = \frac{\varepsilon_1^{02}}{1 + \varepsilon_1^{02}} b_1 = \frac{E_1^{02}}{E_1^{P2} + E_1^{02}} b_1$$

The ratio of the observation error variance to the total error variance determines how large a fraction of the observed departure is analysed.

The analysis at any other position is the product of the analysis at the observation point and the forecast error correlation μ_{k1} between the observation position and the analysis position.

5.4 Prediction error statistics

The long term statistical behaviour of the atmosphere or the forecast model enters through the terms $\langle \alpha_i^P \alpha_j^P \rangle$. In principle these can be functions of several variables

$$\langle \alpha_i^P \alpha_j^P \rangle = F \text{ (variable i, variable j, } x_i, x_j, y_i, y_j, p_i, p_j, t_i, t_j) \qquad (5.4.1)$$

Obviously, F would in its complete form be a very complicated function and its determination from observations almost impossible. Furthermore, F must be positive definite in its domain of use. Some simplifying assumptions are usually made to put F in a mathematically more tractable form. We start by assuming that F is separable and can be written as a product of three positive-definite functions

$$F = H\,(\text{var}_i, \text{var}_j, x_i, x_j, y_i, y_j) \cdot V\,(\text{var}_i, \text{var}_j, p_i, p_j) \cdot T\,(\text{var}_i, \text{var}_j, t_i, t_j) \qquad (5.4.2)$$

Most analysis schemes select observations from a narrow time window around the analysis time. As a consequence the temporal correlation is usually ignored, i.e. $T \equiv 1$ for all time separations $|t_i - t_j|$.

Physical relationships between analysis variables lead to additional simplifications of the correlation functions. Later in this section we will impose geostrophic, hydrostatic and non-divergent constraints on the correlations.

Depending on the particular analysis system the vertical correlation is specified either as a discrete matrix (ECMWF: Lorenc, 1981) or by a continuous function (NMC: Bergman, 1979).

The horizontal correlations have to be modelled by continuous functions due to non-uniform data distribution. In 5.4.1 we will derive expressions for all combinations of height and wind correlations and then relate them geostrophically. In section 5.4.2 we derive hydrostatically coupled vertical correlations.

5.4.1 Horizontal forecast error correlations

Throughout the derivation of the horizontal covariances we assume that the variances are locally constant and that the correlation H can be written in terms of an isotropic and homogeneous function Π as follows

$$\langle a_i \; b_j \rangle = E_a E_b \, H(\Pi, \Pi', \Pi''; \; x_i - x_j, \; y_i - y_j) \qquad (5.4.3)$$

$\Pi^{(n)}$ depends on the separation distance r_{ij} only.

Three separate groups of covariances can be distinguished:

1) z - z
2) u - u, u - v, v - v
3) z - u, z - v

The height error covariance is expressed by

$$\langle z_i \; z_j \rangle = E_z^2 \, \Pi_{zz}(r_{ij}) \qquad (5.4.4)$$

Next, we split the wind in its rotational and divergent part and derive the u - v covariance

$$\begin{aligned}
\langle u_i \; v_j \rangle &= \langle (-\frac{\partial}{\partial y_i} \psi_i + \frac{\partial}{\partial x_i} \chi_i)(\frac{\partial}{\partial x_j} \psi_j + \frac{\partial}{\partial y_j} \chi_j) \rangle \\
&= -\frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j} \langle \psi_i \; \psi_j \rangle + \frac{\partial}{\partial x_i}\frac{\partial}{\partial y_j} \langle \chi_i \; \chi_j \rangle \\
&\quad - \frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j} \langle \psi_i \; \chi_j \rangle + \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} \langle \chi_i \; \psi_j \rangle \\
&= -E_\psi^2 \frac{\partial}{\partial y_i}\frac{\partial}{\partial x_j} \Pi_{\psi\psi} + E_\chi^2 \frac{\partial}{\partial x_i}\frac{\partial}{\partial y_j} \Pi_{\chi\chi} \\
&\quad + E_\psi E_\chi \left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} - \frac{\partial}{\partial y_i}\frac{\partial}{\partial y_j}\right) \Pi_{\psi\chi}
\end{aligned} \qquad (5.4.5)$$

As an example of a covariance from the third group we take the z - v error structure

$$\begin{aligned}
\langle z_i \; v_j \rangle &= \langle z_i (\frac{\partial}{\partial x_j} \psi_j + \frac{\partial}{\partial y_i} \chi_i) \rangle \\
&= \frac{\partial}{\partial x_i} \langle z_i \psi_j \rangle + \frac{\partial}{\partial y_j} \langle z_i \; \chi_j \rangle \\
&= E_z E_\psi \frac{\partial}{\partial x_j} \Pi_{z\psi} + E_z E_\chi \frac{\partial}{\partial y_j} \Pi_{z\chi}
\end{aligned} \qquad (5.4.6)$$

Before specifying the function Π we will assume that the wind error is totally in the rotational part, i.e. $E_\chi = 0$. The differentiation operator can be written as follows

$$\frac{\partial}{\partial x_i} = \frac{\partial r}{\partial x_i}\frac{\partial}{\partial r} = \frac{x_i - x_j}{r}\frac{\partial}{\partial r} \qquad (5.4.7)$$

in which $r = r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$

With (5.4.7) and E_χ=0 the u - v covariance (5.4.5) simplifies to

$$\langle u_i\, v_j\rangle = -E_\psi^2 \frac{(y_i - y_j)}{r}\frac{\partial}{\partial r}\left\{\frac{(x_i - x_j)}{r}\frac{\partial}{\partial r}\Pi_{\psi\psi}\right\}$$

$$= E_\psi^2 \frac{(x_i - x_j)(y_i - y_j)}{r^2}\left\{\Pi''_{\psi\psi} - \frac{1}{r}\Pi'_{\psi\psi}\right\} \qquad (5.4.8)$$

Similarly, for the z - v covariance

$$\langle z_i\, v_j\rangle = -E_z E_\psi \frac{(x_i - x_j)}{r}\Pi'_{z\psi} \qquad (5.4.9)$$

(5.4.4), (5.4.8) and (5.4.9) cover all combinations of height and wind correlations.

The necessary and sufficient conditions for geostrophy between height and wind errors are

1) $\Pi'_{zz} \equiv \Pi'_{\psi\psi} \equiv \text{sign}(f)\,\Pi'_{z\psi}$
This means that the covariances $\langle z_i\, z_j\rangle$, $\langle z_i\, \psi_j\rangle$ and $\langle \psi_i\, \psi_j\rangle$ may contain large scale modes with no associated geostrophic wind.

2) The variances of z and ψ must be geostrophically related.

The final step is to specify Π. Several candidate functions have been considered and used:

$$e^{-ar},\ e^{-ar^2},\ \cos(ar)e^{-br},\ J_0(ar),\ \left[\cos(ar) + \frac{c}{a}\sin(ar)\right]e^{-cr}$$

The function must be twice differentiable (cf. 5.4.8) and its derivative must possess certain properties at the origin (Julian and Thiébaux, 1975).

In the ECMWF analysis scheme it is assumed that the horizontal structure of the forecast error can be represented by a Gaussian type function

$$\Pi(r) = \exp(-0.5\frac{r^2}{b^2}) \qquad (5.4.10)$$

in which b is a tunable scaling parameter (appropriate values of b are 500 - 1000 km). Differentiation of (5.4.10) with respect to r gives

$$\Pi' = -\frac{r}{b^2}\Pi$$
$$\Pi'' = \left(\frac{r^2}{b^4} - \frac{1}{b^2}\right)\Pi \qquad (5.4.11)$$

Substitution of (5.4.10) and (5.4.11) into (5.4.4) and (5.4.8) with extension to all possible wind covariances gives

$$\langle z_i\, z_j\rangle = E_z^2\,\Pi(r)$$
$$\langle u_i\, u_j\rangle = \frac{E_\psi^2}{b^2}\left(1 - \frac{(y_i - y_j)^2}{b^2}\right)\Pi \qquad (5.4.12)$$
$$\langle u_i\, v_j\rangle = \langle v_i\, u_j\rangle = \frac{E_\psi^2}{b^2}\frac{x_i - x_j}{b}\frac{y_i - y_j}{b}\Pi$$
$$\langle v_i\, v_j\rangle = \frac{E_\psi^2}{b^2}\left(1 - \frac{(x_i - x_j)^2}{b^2}\right)\Pi$$

The variance of the wind field is

$$\lim_{j \to i} \langle u_i \, u_j \rangle = \frac{k_\psi^2}{b^2} = \frac{\langle \psi^2 \rangle}{b^2} = \frac{q^2}{f^2} \frac{\langle z^2 \rangle}{b^2} = \frac{q^2}{f^2} \frac{E_z^2}{b^2} \qquad (5.4.13)$$

or $E_\psi = \frac{q}{f} E_z$. The scale parameter b defines the ratio between the height and the wind error.

The cross correlations (z - u and z - v) can be used to relax the geostrophic constraint in the following way

$$\Pi_{z\psi} = G(\phi)\Pi \qquad (5.4.14)$$

where G is a function of latitude and $|G(\phi)| \leq 1$. G changes sign at the equator.

The height-wind correlations can now be written as follows

$$\langle z_i \, u_j \rangle = - \left| \frac{q}{f} \right| \frac{E_z^2}{b} \frac{y_i - y_j}{b} \Pi \cdot G(\phi) \qquad (5.4.15)$$

$$\langle z_i \, v_j \rangle = \left| \frac{q}{f} \right| \frac{E_z^2}{b} \frac{x_i - x_j}{b} \Pi \cdot G(\phi)$$

All 9 possible height and wind correlations are shown in Fig. 5.2 for b = 600 km and $G(\phi) = 1$.

5.4.2 Vertical forecast error correlations

In a three-dimensional scheme vertical correlation matrices (or continuous functions) must be specified for Z and $\underline{V}$. Z - T and T - T or $\underline{V} - \underline{V}_T$ and $\underline{V}_T - \underline{V}_T$ can be derived hydrostatically from Z - Z or $\underline{V} - \underline{V}$. Multiplication of the vertical height error covariance matrix V_{zz} by a differentiation operator

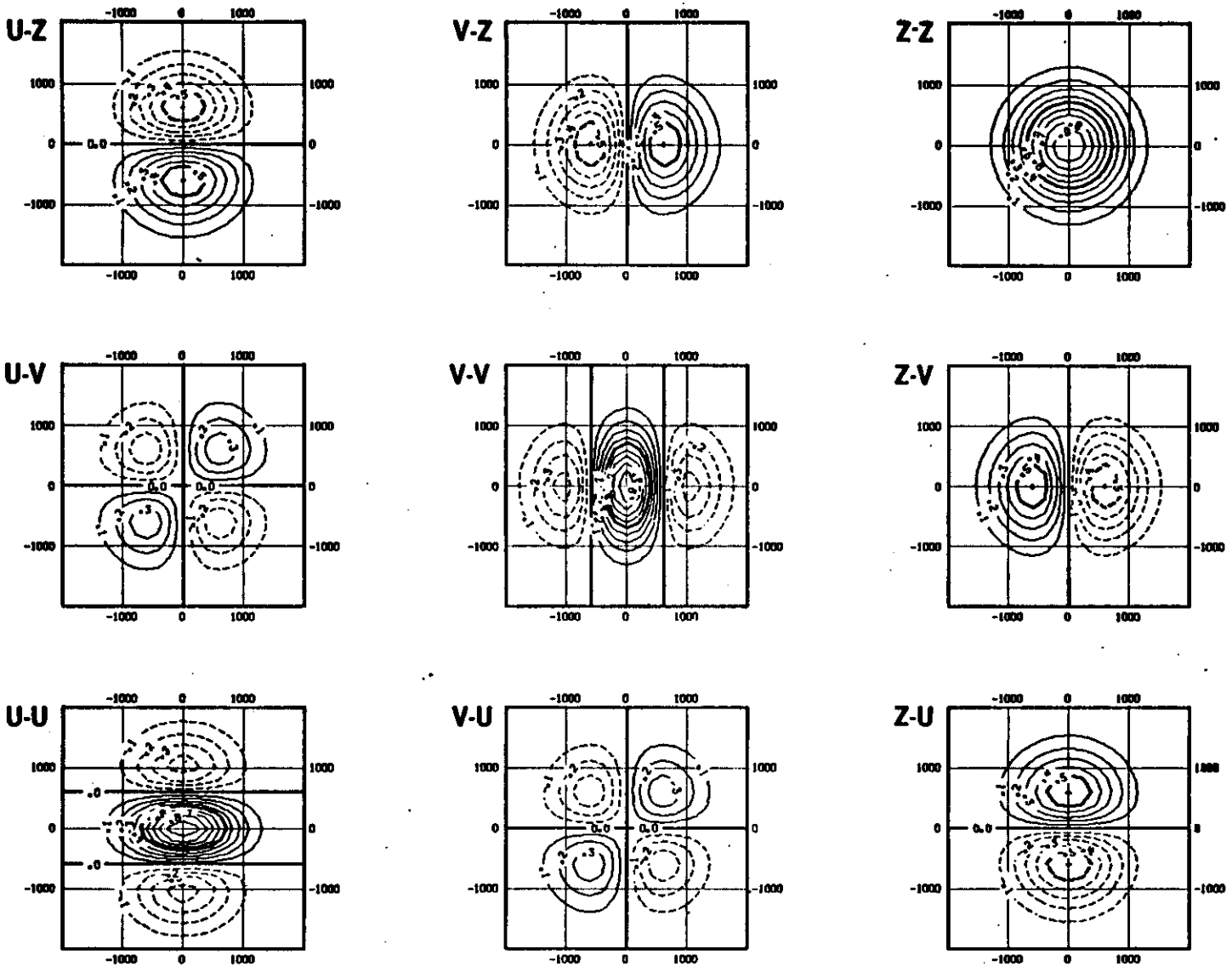


Fig. 5.2 Height and wind correlations based upon $\exp(-0.5r^2/b^2)$ as a function of separation (km). b is 600 km.

$$L_{ZT} = \begin{bmatrix} -\alpha_1 & \alpha_1 & & & \\ & \ddots & \ddots & & 0 \\ & & \ddots & \ddots & \\ 0 & & & -\alpha_{n_o-1} & \alpha_{n_o-1} \end{bmatrix} \qquad (5.4.16)$$

gives V_{ZT}. In (5.4.16) $\alpha_i = \frac{g}{R}(\ell n p_i - \ell n p_{i+1})^{-1}$.

Postmultiplication of V_{ZT} by $L_{TZ} = L_{ZT}^T$ results in the temperature covariance matrix V_{TT}.

The thermal wind covariance matrix $\underline{V}_T - \underline{V}_T$ can be calculated by an identical procedure applied to the vertical wind covariance matrix.

6. SPECTRAL RESPONSE OF AN OPTIMAL INTERPOLATION ANALYSIS

Introduction

In this Chapter we consider at a simple level a number of related questions - how does an OI analysis respond to data, how faithfully does it draw to it and how do the constraints of geostrophy and non-divergence affect the analysis. We begin in Sect.6.1 with a spectral theory which is used throughout the discussion. In Sect.6.2 we present some very simple analytical examples of the analysis of three observations at the observation points, which show some interesting aspects of the theory. Finally in Sect.6.3 we present some numerical results for slightly more general cases to verify some of the suggestions made in Section 6.2.

6.1 Spectral response of the O.I. analysis

In this section we develop some theory that will be useful in discussing the resolution of an O.I. analysis.

As shown in Eqn.5.2.9

$$B_k^i = \underset{\sim}{P}_k^T \underset{\sim}{M}^{-1} \underset{\sim}{B} \qquad (6.1)$$

where B_k^i is the analysed (or interpolated) increment at point k (normalised).

$\underset{\sim}{B}$ is the vector of normalised observation deviations

$\underset{\sim}{M} = \underset{\sim}{P} + \underset{\sim}{Q}$

$\underset{\sim}{P}$ is the prediction error correlation matrix defined by the observation positions and variables.

$\underset{\sim}{Q}$ is the normalised observation error correlation matrix

$\underset{\sim}{P}_k$ is the vector of correlations between the observed variables at the observation positions and the analysed variable at position k.

In order to discuss the resolution of the O.I. analysis it is convenient to specialise the interpolation formula (6.1) to the case where k is an

observation point and B_k^i corresponds to an analysed variable at that point. In other words we consider the problem of analysing the observations at the observation points. If it helps to fix ideas one can suppose the observations to be equally spaced observations of a single variable, say height, although what follows does not depend on such an assumption.

Given that B_k^i corresponds to an analysed variable at an observation position, we may write:

$$\underline{\beta}^i = \underline{P}^T \underline{M}^{-1} \underline{\beta} \qquad (6.2)$$

where $\underline{\beta}$ is the vector of observations and $\underline{\beta}^i$ is the vector of analysed deviations. The operator

$$\underline{P}^T \underline{M}^{-1} \qquad (6.3)$$

is linear. We can say rather a lot about the response properties of the analysis system by examining the eigenvalues and eigenvectors of the analysis operator

$$\underline{P}^T \underline{M}^{-1}$$

To make the mathematics tractable we make the simplifying, and not very strong, assumption, that

$$\underline{Q} = \sigma^2 \underline{I} \qquad (6.4)$$

i.e. that the observation errors are random and uncorrelated with equal (normalised) standard deviation σ.

Since $\underline{P}$, the prediction error correlation matrix, is symmetric, positive definite, we may write

$$\underline{P} = \underline{E}[\lambda]\underline{E}^T \qquad (6.5)$$

where $\underline{E}$ is the matrix of orthogonal normalised eigenvectors and the matrix $[\lambda]$ is diagonal and has the eigenvalues of $\underline{P}$ on the main diagonal. If there are N observations then, since $\underline{P}$ is a correlation matrix with diagonal elements 1, a standard result gives us that

$$\mathrm{tr}(\underline{P}) = N = \sum_{i=1}^{N} \lambda_i \qquad (6.6)$$

and $\lambda_i > 0$ for all i

Because of our simplifying assumption on $\underline{Q}$ we may write

$$\underline{M} = \underline{P} + \underline{Q} = \underline{P} + \sigma^2 \underline{I} = \underline{E}[\lambda+\sigma^2]\underline{E}^T$$

Thus the effect of adding the observational error matrix to $\underline{P}$ is to increase all the eigenvalues by the value σ^2, without altering the eigenvectors. It follows that

$$\underline{M}^{-1} = \underline{E}[1/(\lambda+\sigma^2)]\underline{E}^T \qquad (6.7)$$

and finally that

$$\underline{P}^T \underline{M}^{-1} = \underline{E}[\lambda/(\lambda+\sigma^2)]\underline{E}^T \qquad (6.8)$$

This is a very powerful result as can be shown if we write

$$\underline{\beta} = \underline{E}\,\underline{\beta}' \quad , \quad \underline{\beta}^i = \underline{E}\,\underline{\beta}^{i\prime} \qquad (6.9)$$

and substitute 6.8, 6.9 in 6.2

$$\underline{E}\,\underline{\beta}^{i\prime} = \underline{E}[\lambda/\lambda+\sigma^2]\underline{\beta}' \qquad (6.10)$$

or

$$\underline{\beta}^{i\prime} = [\lambda/\lambda+\sigma^2]\underline{\beta}' \qquad (6.11)$$

The eigenvectors $\underline{E}$ will, in simple situations (equally spaced observations of a single variable) have a natural ordering of scales associated with them. The result (6.11) then gives a clear expression of the resolving power of the analysis. $\underline{\beta}'$, $\underline{\beta}^{i\prime}$ represent expansion coefficients of the data and the analysis increments in terms of a complete orthogonal basis. For well resolved modes, $\lambda_i \gg \sigma^2$, we have $\beta_k^{i\prime} = \beta_k'$. In other words the analysis draws very closely to the data in these components. For poorly resolved modes, $\lambda_i \ll \sigma^2$, we have $\beta_k^{i\prime} \ll \beta_k'$, ie. these components of the data are heavily damped in the analysis.

Note that

$$\frac{\lambda_i}{\lambda_i + \sigma^2} < 1 \text{ for all } i$$

so that the O.I. operator damps every component in the data. This is because it is taking some information [in ratio = $\frac{\sigma^2}{\lambda_i + \sigma^2}$] from the first guess.

The spread of the eigenvalues λ_i determines the response of the analysis. This spread is governed by the position and type of variable observed. Some simple limiting cases are of interest.

Limit 1 Widely spaced observations

If all the observations are so far apart that they are quite independent, then

$$\underset{\sim}{E} = \underset{\sim}{I},$$

and

$$\lambda_i = 1, \ \underset{\sim}{E} = \underset{\sim}{I}$$

so that

$$B_k^i = \frac{1}{1+\sigma^2} B_k$$

In other words the analysis increment at the observation points is $\frac{1}{1+\sigma^2}$ the observed deviation. This is the optimal combination of two independent estimates of the true value, one provided by the first guess (with normalised error variance 1) and the other provided by the observation (with normalised error variance σ^2). This interpretation is based on the simple result that if we have two estimators of x, viz., x_1, x_2 with error variances σ_1^2 and σ_2^2 then the optimal estimator of x is

$$\tilde{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2$$

Limit 2 Closely spaced observations

At the other extreme we have the case where all the observations of the single variable are so close together that all the entries of $\underset{\sim}{P}$ are 1.

Such a matrix is clearly of rank 1, so that it has only a single non-zero eigenvalue, $\lambda_1 = N$ with eigenvector $(1,1\ldots1)/\sqrt{N}$

follows that $\beta_1^i = \frac{N}{N+\sigma^2} \beta_1$

and

$$B_1^i = \frac{1}{1+\sigma^2/N} \left(\frac{\Sigma B_i}{N}\right) .$$

Then the analysed value is $\frac{1}{1+\sigma^2/N}$ times the mean value of the observations. Physically this corresponds to the optimal combination of the first guess with variance 1 and a single super-observation with error variance σ^2/N. All other structure in the data is ignored.

In order to explore our theoretical results, it is useful to consider some simple examples, bearing in mind the two limiting cases we have just discussed. It should also be borne in mind that if the observations are exact (σ^2=o) then all the data are drawn for exactly.

6.2 Some simple three point analysis problems

In this Section we consider a few examples of some very simple analysis problems, involving the analysis of three observations, at the observation points. These examples are very simple, but are instructive so that it is worthwhile working through them in detail.

There are a few preliminaries which we need to refer to. We will be using two different functions for the $\langle\phi\phi\rangle$ structure function, the Gaussian and the Gauss-Markovian function of second order. We shall assume the height stream-function correlation is 1 and we shall assume that the reader has read and understood Chapter 5 of these notes. For convenience we summarise the properties of the Gaussian and Gauss-Markovian in 1-dimension using $\xi = \Delta x/b$

Structure Function	Gaussian	Gauss-Markovian
$\langle\phi\phi\rangle$	$e^{-1/2\xi^2}$	$(1+\lvert\xi\rvert)\, e^{-\lvert\xi\rvert}$
$\langle\phi V\rangle$	$-\xi e^{-1/2\xi^2}$	$\pm\lvert\xi\rvert\, e^{-\lvert\xi\rvert}$
$\langle VV\rangle$	$-(\xi^2-1)e^{-1/2\xi^2}$	$-(\lvert\xi\rvert-1)e^{-\lvert\xi\rvert}$

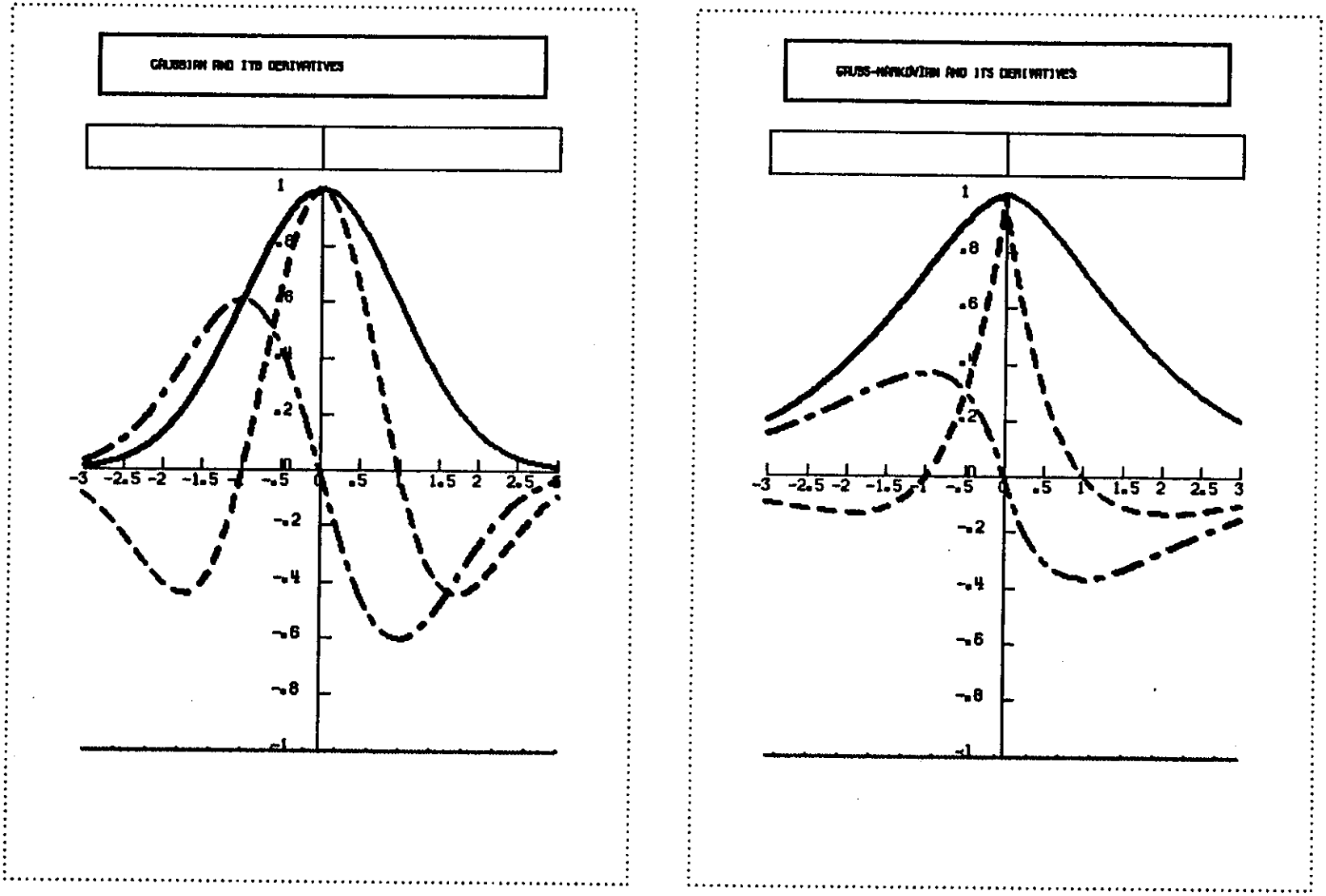


Figure 6.1 Plots of the function (————), it's derivative (– – – –) and it's second derivative (— - — - — -) for the two one dimensional structure functions considered; the Gaussian on the left, and the second order Gauss-Markovian on the left.

Plots of each of these functions are shown in Fig.6.1. We note that the functions are qualitatively similar but differ, for example, in the depth of the negative lobes of the <VV> correlation, which occur because of our assumption of non-divergence.

EXAMPLE 1 Three heights

We consider the problem of analysing three equally spaced collinear height observations

ϕ_1 ϕ_3 ϕ_2

+Δx+ +Δx+

The $\underline{P}$ matrix may be written

$$\begin{bmatrix} 1 & p & q \\ p & 1 & q \\ q & q & 1 \end{bmatrix}$$

where $p = \langle\phi_1\phi_2\rangle$

$q = \langle\phi_1\phi_3\rangle = \langle\phi_2\phi_3\rangle$

We have chosen the numbering for the grid points in order to make comparisons later when we replace the ϕ_3 observation by a v_3 observation.

The equation for the eigenvalues λ is

$$(1-\lambda)\{(1-\lambda)^2 - p^2\} - 2q^2\{(1-\lambda) - p\} = 0$$

For which the roots are

$$\lambda_1 = 1 + p/2 + r$$

$$\lambda_2 = 1 - p$$

$$\lambda_3 = 1 + p/2 - r$$

where $r = \{(p/2)^2) + 2q^2\}^{1/2}$

Since we require that all the eigenvalues shall be non-negative then we must have $\lambda_3 > 0$, i.e.

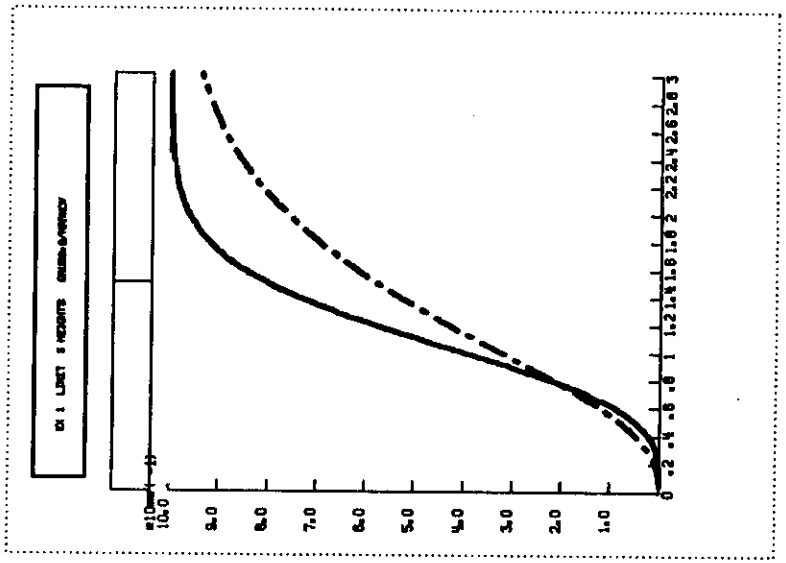


Figure 6.2 Plots of the degree to which the condition for positive-definiteness is satisfied in the first example (three heights). The solid line is for the Caussian, the other for the Gauss-Markovian.

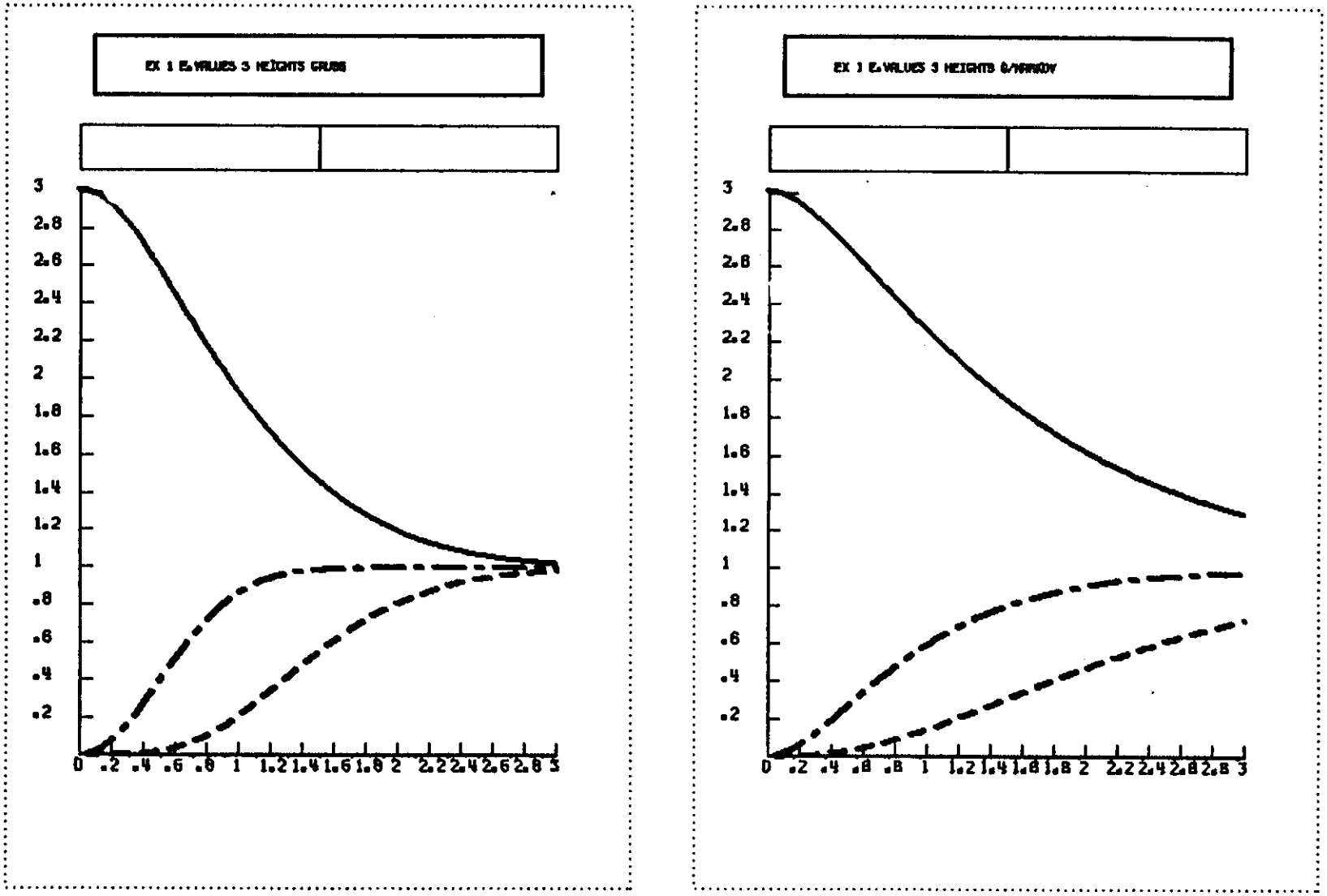


Figure 6.3 Plots of the dependence of the three eigenvalues of the first example (three heights) on the normalised spatial separation parameter for the Gaussian (left) and the Gauss Markovian (right)

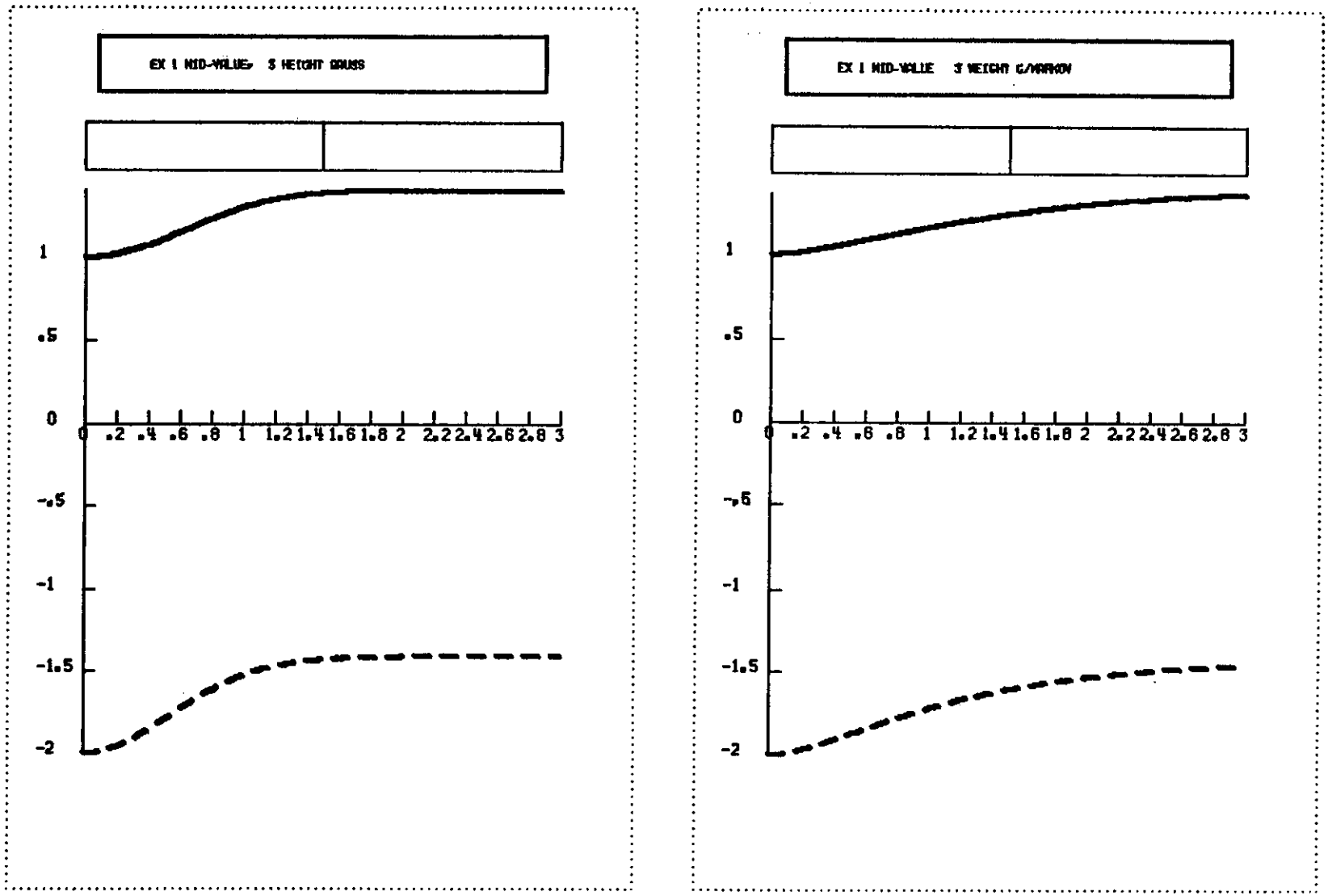


Figure 6.4 Plots of the variation with normalised observation separation of the values of the non-zero mid-point elements in the eigenvectors of the first example (three heights) The solid line corresponds to the mean-value eigenvector, while the other corresponds to the two grid-length wave; results for the Gaussian on the left, Gauss-Markovian on the left.

$$f(\xi) = 1 + p - 2q^2 > o$$

This is readily satisfied by both structure functions as shown in Fig.6.2. Near the origin $f(\xi)$ for the Gaussian is smaller than that for the Gauss-Markovian.

In order to study the eigenvalues we follow the usual procedure of solving for the eigenvalues, discarding one equation, specifying the corresponding variable (in our case ϕ_3) and then solving the remaining equations.

The first two equations then become

$$(1-\lambda)\phi_1 + p\phi_2 = -q\phi_3$$

$$p\phi_1 + (1-\lambda)\phi_2 = -q\phi_3.$$

These can only be solved for non zero $(q\phi_3)$ if

$$(1-\lambda)^2 - p^2 = (1-p-\lambda)(1-p+\lambda) \neq o.$$

Thus for $\lambda_2 = 1-p$ we must require $(q\phi_3)=o$. This then gives us that $p(\phi_1+\phi_2)=o$ or $\phi_1 = -\phi_2$. Thus the second eigenvector is given by

$$\lambda_2 = 1-p, \quad e_2 = (1, 0, -1)$$

and clearly corresponds to a linear trend in the data.

For the other two eigenvectors, we can solve the eigenvector equations by Cramer's rule and get

$$\phi_1 = \phi_2 = -\frac{q\phi_3}{1+p-\lambda}$$

Thus the eigenvectors are

$$\lambda_1 = 1 + p/2 + r \qquad e_1 = \left(1, \frac{(r-p/2)}{q}, 1\right)$$

$$\lambda_2 = 1 - p \qquad e_2 = (1, o, -1)$$

$$\lambda_3 = 1 + p/2 - r \qquad e_3 = \left(1, -\frac{(r+p/2)}{q}, 1\right).$$

Since $r = \sqrt{\left(\frac{p}{2}\right)^2 + 2q^2}$ it follows that all three eigenvectors are orthogonal.

Fig.6.3 shows the plots of λ_1, λ_2, λ_3 for $0 \leq \frac{\Delta x}{b} \leq 3$ for the Gaussian (6.3a)

and the Markovian (6.3b). Fig.6.4 shows the behaviour of the elements $\phi_3(\lambda_1)$, $\phi_3(\lambda_3)$ for the Gaussian and the Gauss-Markovian.

The above expressions for the eigenvectors show that the characterisation of e_1 as a mean value and e_3 as a two grid wave is valid for all values of $\Delta x/b$ (this is not the case in one of the later examples).

Secondly for $\Delta x/b \leq 0.8$, the Gaussian gives a better analysis (λ_2 larger) for the linear trend than does the Gauss-Markovian, while the Gaussian applies heavier smoothing (λ_3 smaller) to the two grid wave, than does the Gauss-Markovian. For the mean value, both structure functions do about equally well.

This example is quite straightforward, and the only surprise perhaps is that only one mode, the gravest, has $\lambda > 1$.

EXAMPLE 2 Three Winds

Our second example is formally identical with our first, but the results are very different.

We consider the problem of three equally spaced wind observations for the component normal to the line of stations

V_1 $\quad +\Delta x+ \quad$ V_3 $\quad +\Delta x+ \quad$ V_2

$$\begin{bmatrix} 1 & p & q \\ p & 1 & q \\ q & q & 1 \end{bmatrix}$$

where $p = \langle V_1 V_2 \rangle$

$q = \langle V_1 V_3 \rangle = \langle V_2 V_3 \rangle$

The condition for positiveness of all the eigenvalues is again

$$f(\xi) = 1 + p - 2q^2 > 0$$

where $p = -\bar{g}(2\xi)$, $q = -\bar{g}(\xi)$ and g is the stream-function structure function. The values of $f(\xi)$ for the Gaussian and the Gauss-Markovian are shown in Fig.6.5. It is clear that the Gaussian gives values for f which are closer to zero over most of the range, and so it will have one eigenvalue smaller than the corresponding Gauss-Markovian eigenvalue.

The roots are exactly as given before

$$\lambda_1 = 1 + p/2 + r$$

$$\lambda_2 = 1 - p$$

$$\lambda_3 = 1 + p/2 - r \quad \text{with } r = [(p/2)^2 + 2q^2]^{1/2}$$

and for $q \neq o$ the eigenvectors are

$$e_1 = (1, \quad \frac{r-p/2}{q}, \quad 1)$$

$$e_2 = (1, \quad o, -1)$$

$$e_3 = (1, \quad -\frac{(r+p/2)}{q}, \quad 1) \text{ as before.}$$

However the categorization of the eigenvectors depends on whether $q \gtrless 0$. (i.e. $\Delta x/b \lessgtr 1$).

Fig.6.6 shows the plots of $\lambda_1, \lambda_2, \lambda_3$ for $0 \leq \Delta x/b \leq 3$. We have deliberately introduced a discontinuity into this plot to indicate that the eigenvector corresponding to the mean-value of the observations has the eigenvalue

$$\lambda = 1 + \frac{p}{2} + r, > 1 \text{ for } \frac{|\Delta x|}{b} < 1 \quad \text{and}$$

$$\lambda = 1 + \frac{p}{2} - r, < 1 \text{ for } \frac{|\Delta x|}{b} > 1 \quad .$$

Similarly the eigenvector corresponding to the two-grid wave has eigenvalue

$$\lambda = 1 + \frac{p}{2} - r, < 1 \text{ for } \frac{|\Delta x|}{b} < 1 \quad \text{and}$$

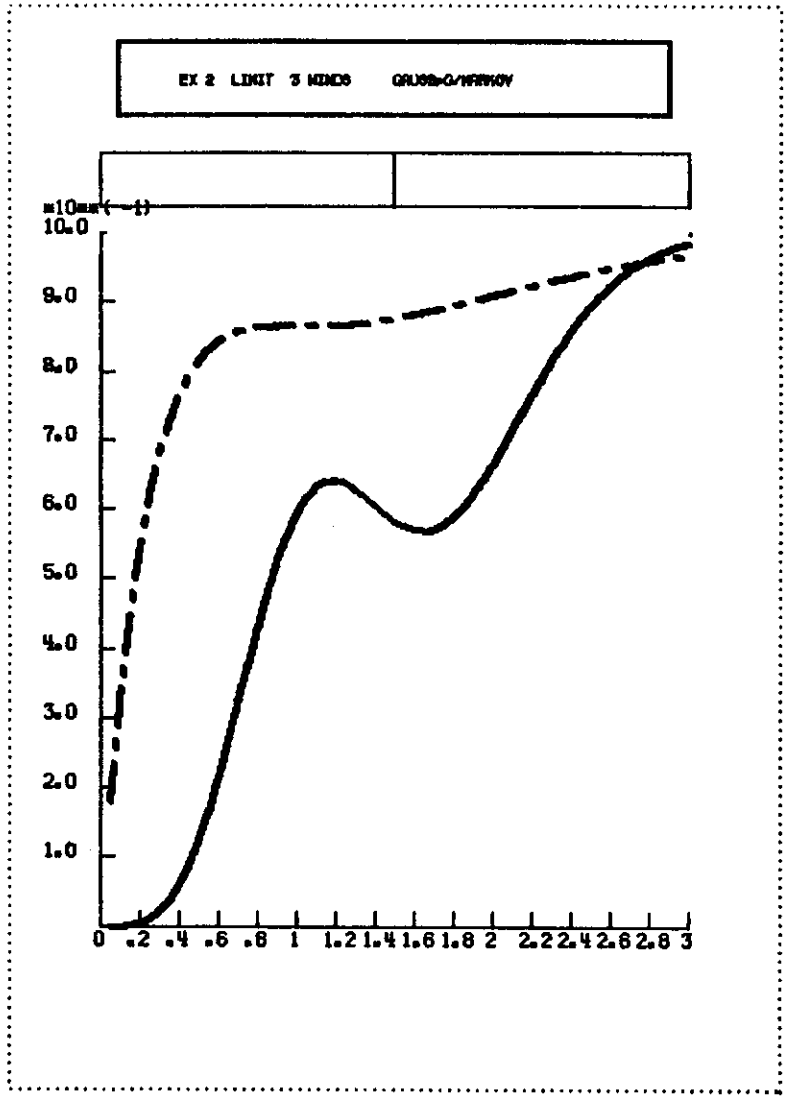


Figure 6.5 Plots of the degree to which the condition for positive-definiteness is satisfied in the second example (three winds). The solid line is for the Gaussian, the other for the Gauss-Markovian.

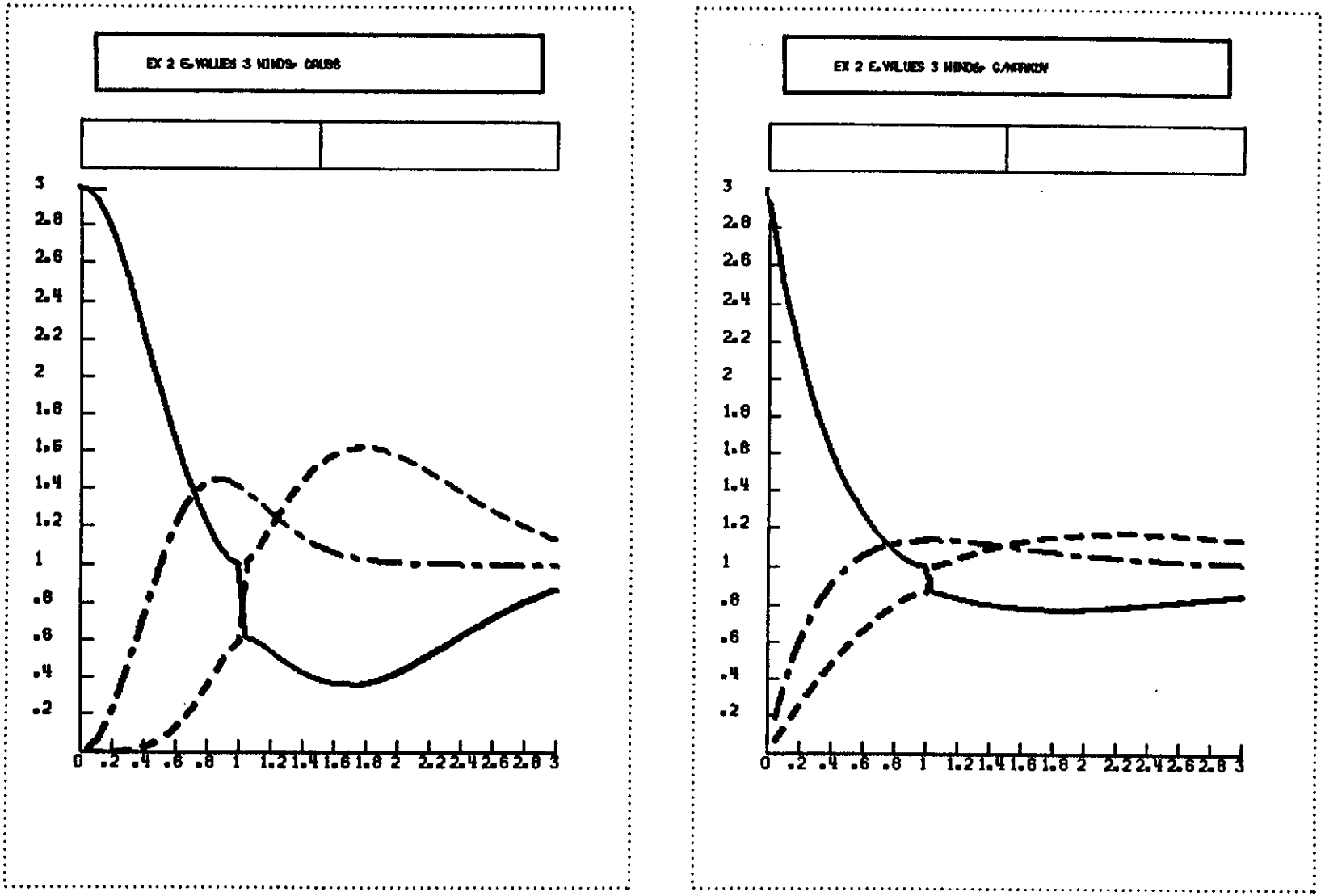


Figure 6.6 Plots of the dependence of the three eigenvalues of the second example (three winds) on the normalised spatial separation parameter for the Gaussian (left) and the Gauss Markovian (right)

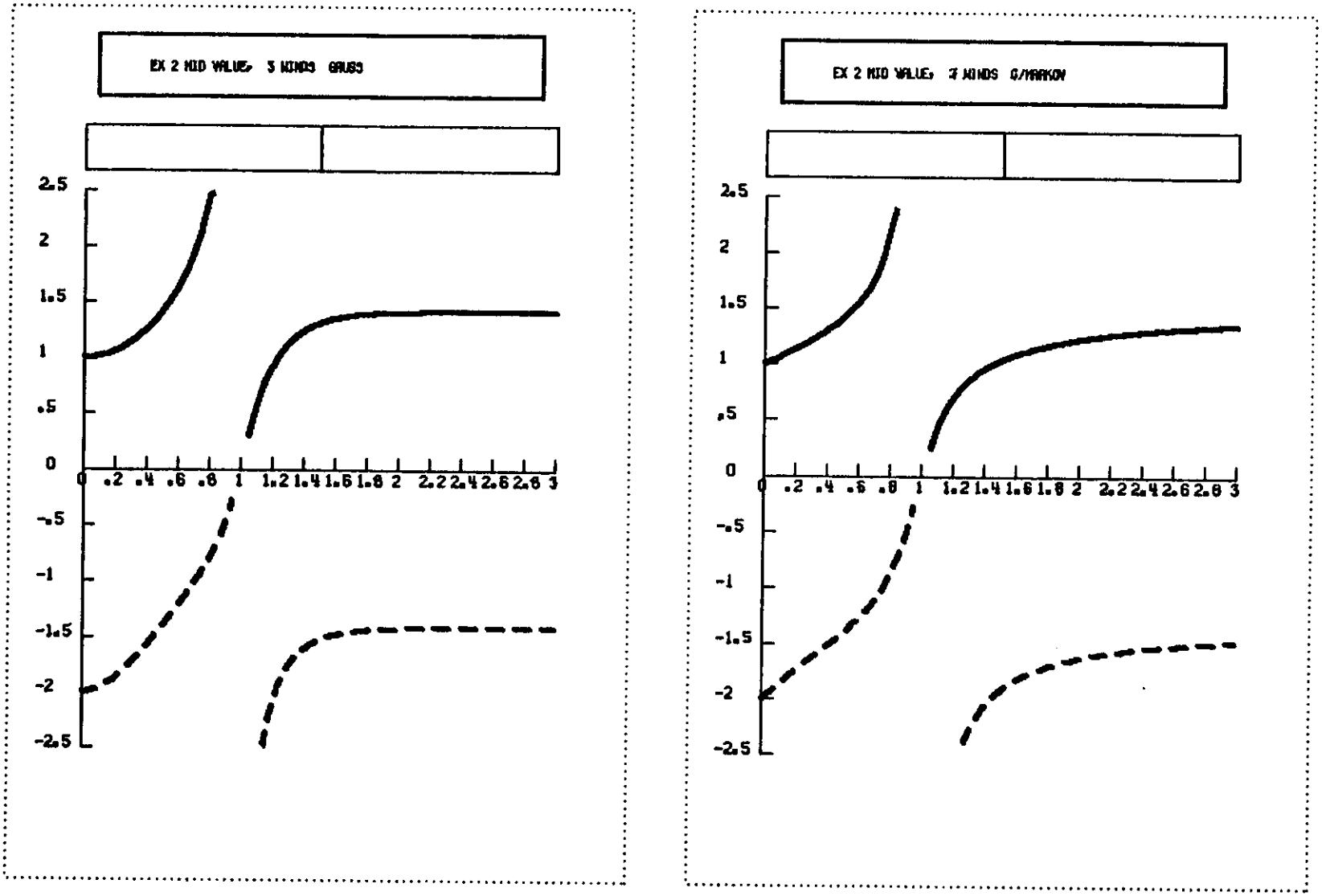


Figure 6.7 Plots of the variation with normalised observation separation of the values of the non-zero mid-point elements in the eigenvectors of the second example (three winds) ; results for the Gaussian on the left, Gauss-Markovian on the left. Further discussion in the text.

$$\lambda = 1 + \frac{p}{2} + r, > 1 \text{ for } \frac{|\Delta x|}{b} > 1$$

In fact for $\frac{\Delta x}{b} \sim \sqrt{3} = 1.73$, the two grid wave has the largest eigenvalue. This is a fairly general result and indicates that some care must be taken in the wind analysis with O.I.

Fig.6.7 plots the value of V_3 in the two extreme eigenvectors. We have not normalised the eigenvectors so that there is a singularity because of the term in 1/q. If we had normalised the vectors then near the singularity the eigenvectors would be $(\varepsilon, -1, \varepsilon), (\varepsilon, +1, \varepsilon)$ $\varepsilon \to o$.

When $\frac{\Delta x}{b} = 1$ (i.e. at the singularity) the eigenvalues and eigenvectors are given by

$$\lambda_1 = 1 \qquad e_1 = (0, 0, 1)$$

$$\lambda_2 = 1 + p \qquad e_2 = (+1, +1, 0)$$

$$\lambda_3 = 1 - p \qquad e_3 = (1, -1, 0)$$

Physically, it is clear what is happening. Given the non-divergent wind-wind structure functions, the analysis expects to find that a northerly wind at one point will be accompanied by a southerly wind at a distance, $\sqrt{3}b$ to the east or west. It will favour such observations in its analysis and will damp observations which run counter to this expectation. In particular, a large-scale slowly varying feature in the observations will be significantly damped. This problem is the subject of current research work.

As expected the Gaussian shows a heavier damping of the mode with smallest (see Fig.6.6) eigenvalue than the Gauss-Markov. Physically this corresponds to the fact that the wind-wind correlations (both positive or negative) are weaker with the Gauss-Markov than with the Gaussian.

EXAMPLE 3 Two heights and a wind

In this problem we envisage a wind observation replacing the central height observation in Example 1.

ϕ_1 $\quad$ V_3 $\quad$ ϕ_2

$\quad$ +Δx+ $\quad$ +Δx+

The matrix in this case is

$$\begin{bmatrix} 1 & p & -q \\ p & 1 & q \\ -q & q & 1 \end{bmatrix}$$

where $p = \langle \phi_1 \phi_2 \rangle$

$q = \langle V_3 \phi_2 \rangle = - \langle V_3 \phi_1 \rangle > 0$

$p = g(2\xi)$, $q = \mu \dot{g}(\xi)$ where g is the height-height (and streamfunction-streamfunction) correlation, $\xi = \frac{\Delta x}{b}$, and μ is the height-stream function correlation which we take to be 1.

$\mu > 0$ in the northern hemisphere and $\mu < 0$ in the southern hemisphere because of the change of sign in the Coriolis parameter.

The minus signs in the matrix occur because of the assumed geostrophy.

The determinantal equation is

$$(1-\lambda)[(1-\lambda)^2 - p^2] - 2q^2(1-\lambda+p) = 0$$

The condition for positive definiteness is

$$f_3\left(\frac{\Delta x}{b}\right) = 1 - p - 2q^2 > 0$$

The values of f in this case are plotted in Fig.6.8. The Gaussian has smaller values of f_3 over the whole range and so it will have one eigenvector smaller and one larger than the Gauss-Markovian. The eigenvalues are given by

$$\lambda_1 = 1 + p$$

$$\lambda_2 = 1 - p/2 + r$$

$$\lambda_3 = 1 - p/2 - r$$

$$r = \sqrt{(\frac{p}{2})^2 + 2q^2}$$

These are plotted on Fig.6.9. We see that there are two modes which are always well analysed while the third mode is significantly damped for $\left|\frac{\Delta x}{b}\right| < 1$. To examine the eigenvectors we proceed as before. The equations we have to solve, for specified V_3 are

$$(1-\lambda)\phi_1 + p\phi_2 = qV_3$$

$$p\phi_1 + (1-\lambda)\phi_2 = -qV_3$$

These have a solution for qV_3 non zero, provided

$$(1-\lambda)^2 - p^2 \neq 0$$

For λ_1, the first eigenvector, $\lambda = 1+p$ so this condition is violated. It follows that for this mode $V_3 = 0$. and then

$p(-\phi_1 + \phi_2) = 0$ which implies that the vector is of the form

ϕ_1 V_3 ϕ_2

(1, 0, 1)

This clearly corresponds to the mean height, and has no wind associated with it. (This latter feature is not a general result, as the "mean height" eigenvector generally does not have identical values at all grid points (cf Example 1). The second and third eigenvalues satisfy the condition

$$(1-\lambda) - p^2 \neq 0$$

and for these we have the result $\phi_1 = -\phi_2 = \frac{q\,V_3}{1-\lambda-p}$

Thus the eigenvectors have the form

		ϕ_1	V_3	ϕ_2	
$\lambda_1 = 1 + p$	e_1	(1	0	1)	
$\lambda_2 = 1 - p/2 + r$	e_2	(-1,	$+\frac{r+p/2}{q}$,	1)	
$\lambda_3 = 1 - p/2 - r$	e_3	(-1,	$-\frac{(r-p/2)}{q}$,	1)	$r = \sqrt{(\frac{p}{2})^2+2q^2}$.

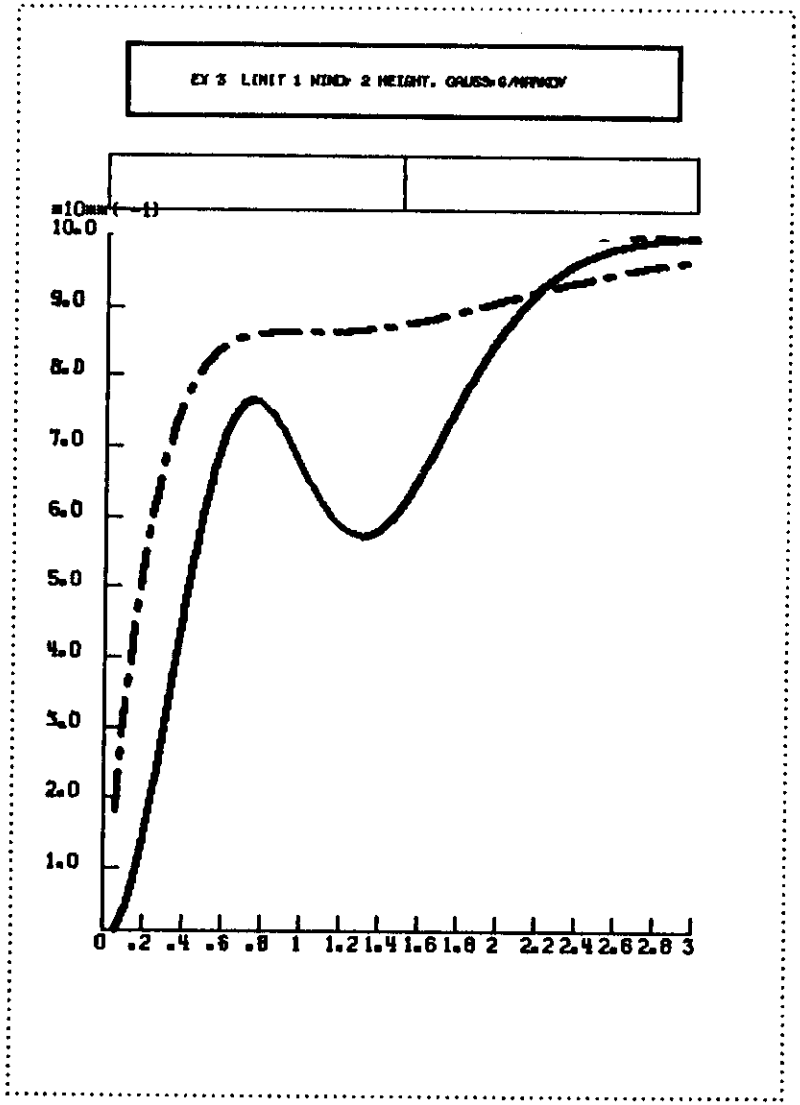


Figure 6.8 Plots of the degree to which the condition for positive-definiteness is satisfied in the third example (one wind and two heights). The solid line is for the Gaussian, the other for the Gauss-Markovian.

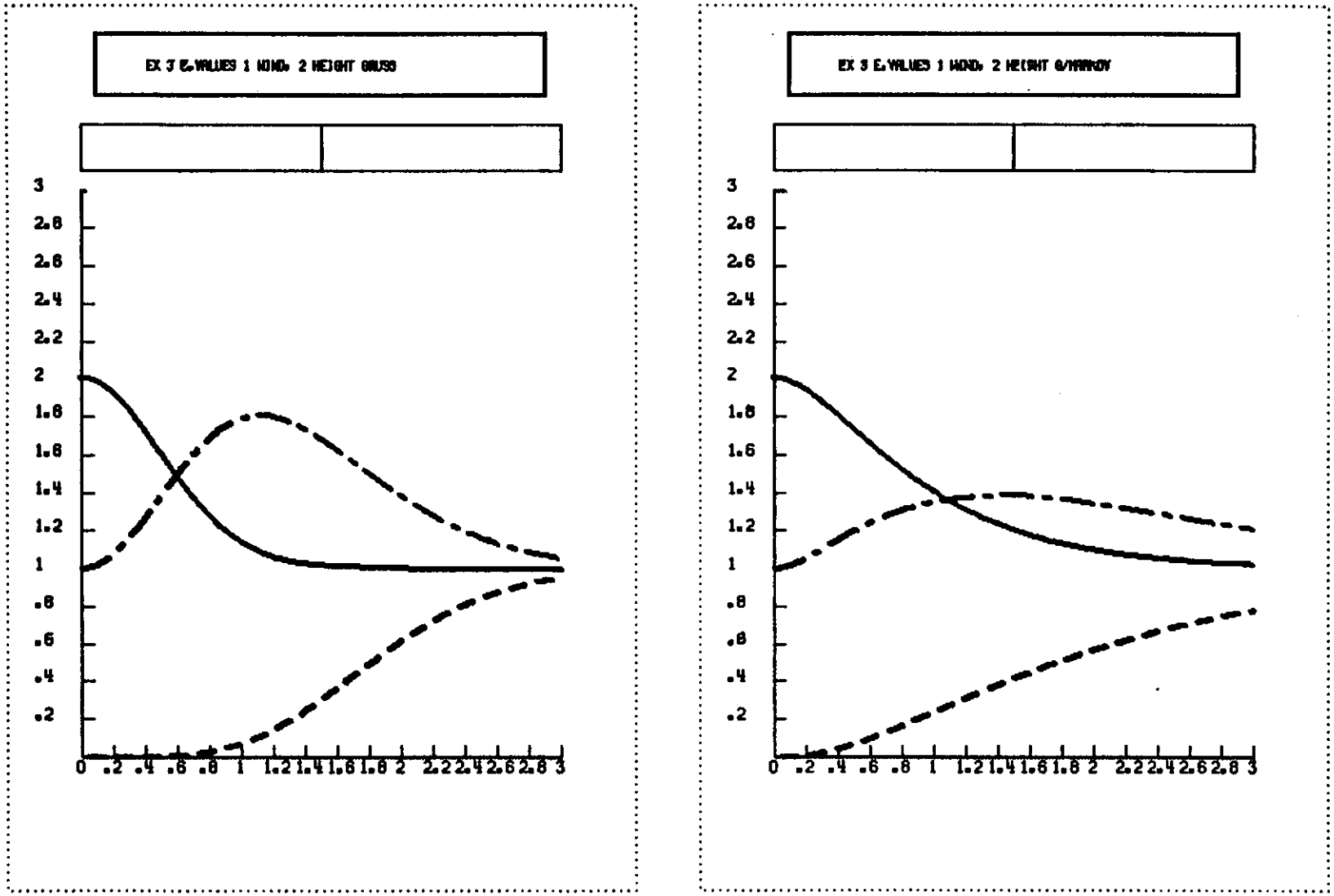


Figure 6.9 Plots of the dependence of the three eigenvalues of the third example (one wind and two heights) on the normalised spatial separation parameter for the Gaussian (left) and the Gauss Markovian (right). The solid line is the mean-height eigenvalue, the smallest eigenvalue is the anti-geostrophic one, while the third is the geostrophic eigenvalue.

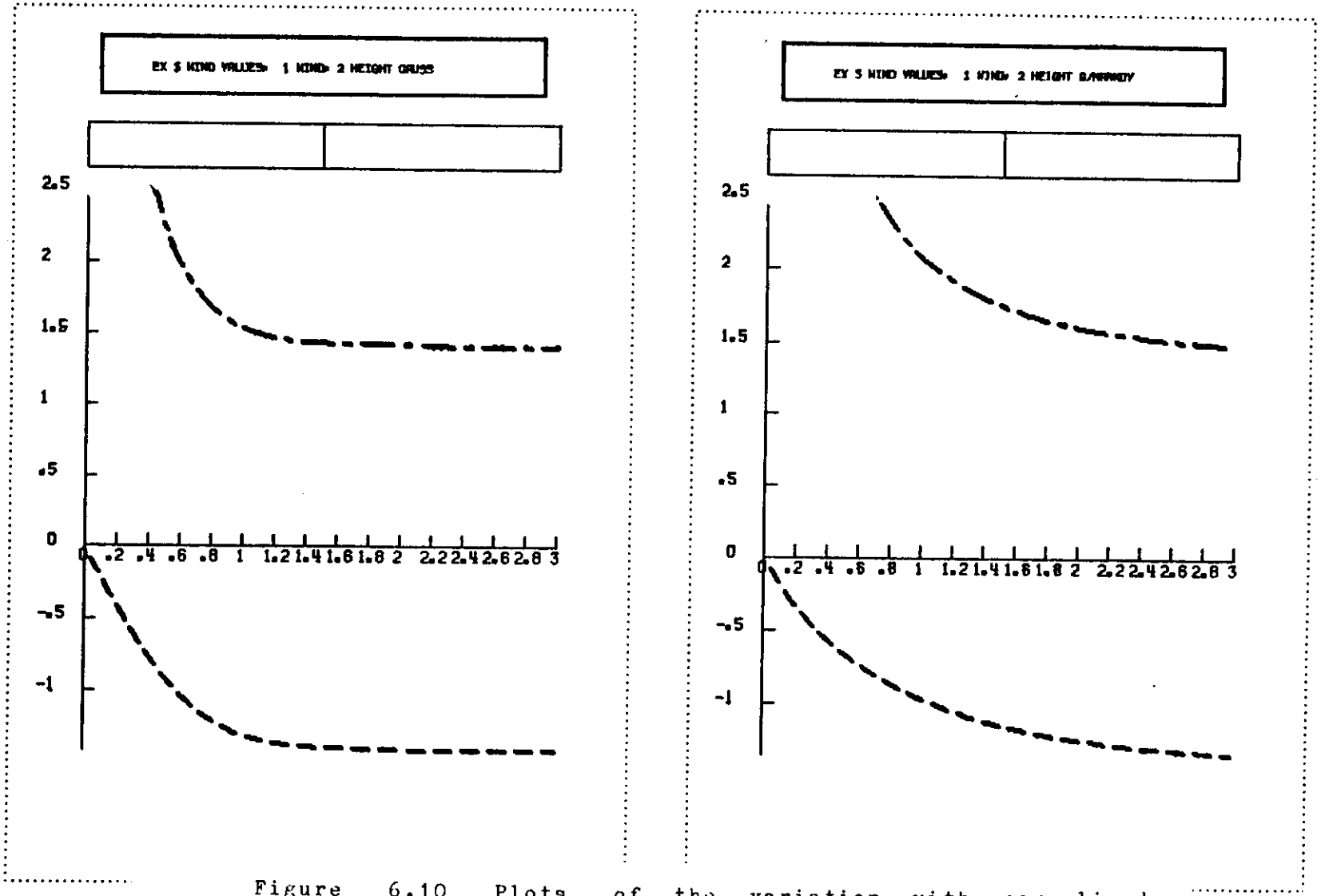
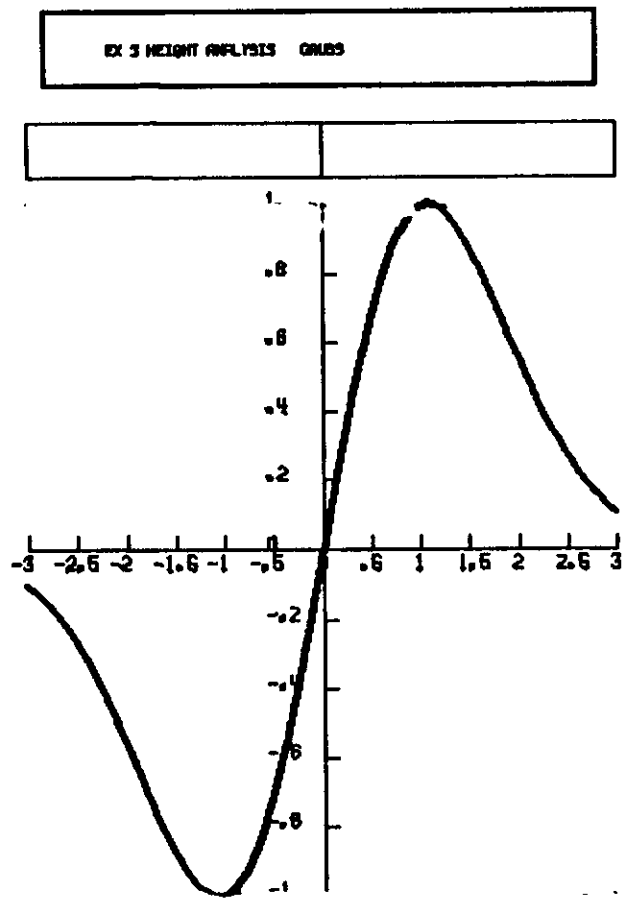


Figure 6.10 Plots of the variation with normalised observation separation of the values of the non-zero wind elements in the eigenvectors of the third example (one wind and two heights); results for the Gaussian on the left, Gauss-Markovian on the left. At small separation a linear trend in the height implies a large geostrophic wind and a vanishing anti-geostrophic wind in the corresponding eigenvectors.



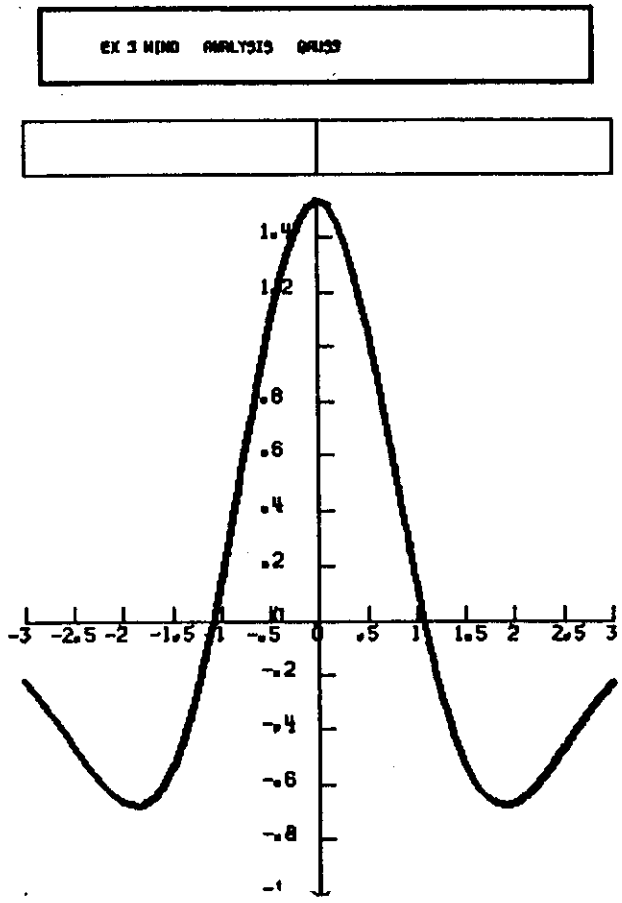


Figure 6.11 Plots of the implied continuous height (right) and wind (left) analyses in example three when the observational data have the same values as the geostrophic eigenfunction. The wind observation is at the origin while the height observations are at unit distance east and west of the wind point. The analysed winds are not precisely zero at the height observation points because of the influence of the reported wind at the mid-point.

The eigenvectors e_2, e_3 have the same height gradient associated with them, but opposite signs for the associated wind. Clearly one of them approximates the geostrophic wind while the other wind must be quite ageostrophic. In order to discuss this point it is convenient to suppose that $\frac{\Delta x}{b}$ is between 1 and 3 where q is not small but p is small. In other words the heights are uncorrelated, but they are correlated with the wind.

	Northern Hemisphere	Southern Hemisphere
$\lambda_2 \sim 1+\sqrt{2}\,q$	$e_2 \sim(-1, \sqrt{2}, 1)$	$(-1, -\sqrt{2}, 1)$
$\lambda_3 \sim 1-\sqrt{2}\,q$	$e_3 \sim(-1, -\sqrt{2}, 1)$	$(-1, \sqrt{2}, 1)$

If we suppose that ϕ_2 is to the east of ϕ_1 then clearly the wind in $\underline{e}_2$ is in the geostrophic direction while in $\underline{e}_3$ it is in the anti-geostrophic direction. We note too that the magnitude of the wind in $\underline{e}_2$ is not the same as the geostrophic wind calculated from ϕ_1 ϕ_2 by a finite difference calculation,

i.e. $$\frac{\phi_1 - \phi_2}{2(\frac{\Delta x}{b})} = 1 \quad \text{when } (\frac{\Delta x}{b}) = 1$$

Fig.6.10 shows that the corresponding value in the eigenfunction is of order 1.5. The reason for the difference is truncation error in the finite difference calculation. The O.I. analysis has an underlying continuous structure, with geostrophy enforced by differentiation. In the present example (cf Fig.6.11) the wind is almost zero at points 1 and 2. It therefore follows that the continuous derivative of height at point 3 (and so the geostrophic wind) has to exceed the finite difference value.

This is a very important point to bear in mind when discussing constraints (e.g. geostrophy, non-divergence) on an O.I. analysis. Since the data are always noisy, they will contain all possible modes. The analysis will damp,

but will not eliminate entirely, those components of the data which do not satisfy the (continuously expressed) constraints. The evaluation of the analysis on a grid will then produce fields in which the dominant eigenvectors will satisfy the constraints for analytic differentiation. If the grid on which the analysis is evaluated is too coarse (e.g. grid size of order b) the gridded analysis will not satisfy the constraints if these are calculated as finite differences on the grid. The present calculation for the geostrophic wind is a clear example of this.

EXAMPLE 4 Two winds and one height

Our fourth example has exactly the same mathematical form as the previous one except that the structure functions change. In the notation of Example 3

V_1 $\qquad$ ϕ_3 $\qquad$ V_2

$\quad$ +Δx+ $\qquad$ +Δx+

$$p = \ddot{g}(2\xi),\ q = \mu\,\dot{g}(\xi) \qquad \xi = \frac{\Delta x}{b},\ \mu = 1$$

The condition for positive definiteness is that

$$f_4(\xi) = 1 - p - 2q^2 > 0 .$$

This is shown in Fig.6.12 for $0 < \xi < 3$. Fig.6.13, 6.14 show the results for eigenvalues and eigenvectors. As we saw in the case of three winds the mean wind eigenvector is significantly damped ($\lambda \sim 0.6$) when the spacing between the winds is of order $\sqrt{3}$. When this happens the negative correlations between the winds are largest. The geostrophic eigenvalue is largest when the spacing between the height and wind observations is of order 1. This gives the largest correlation between these observations. Very closely spaced height and wind observations are essentially treated as being independent. Again the relatively weak wind-wind correlations implied by the Markovian result in the mean wind eigenvector being close to its asymptotic value already when the spacing between the wind observations is of order 1.

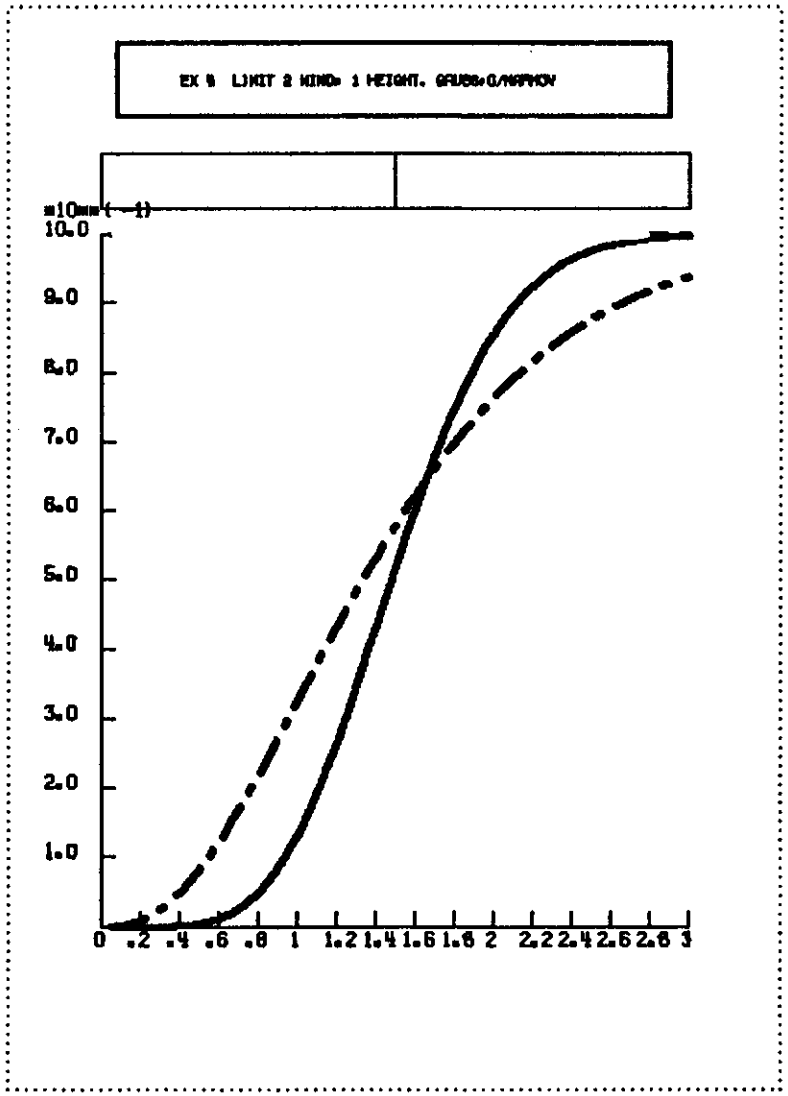


Figure 6.12 Plots of the degree to which the condition for positive-definiteness is satisfied in the fourth example (one height and two winds). The solid line is for the Gaussian, the other for the Gauss-Markovian.

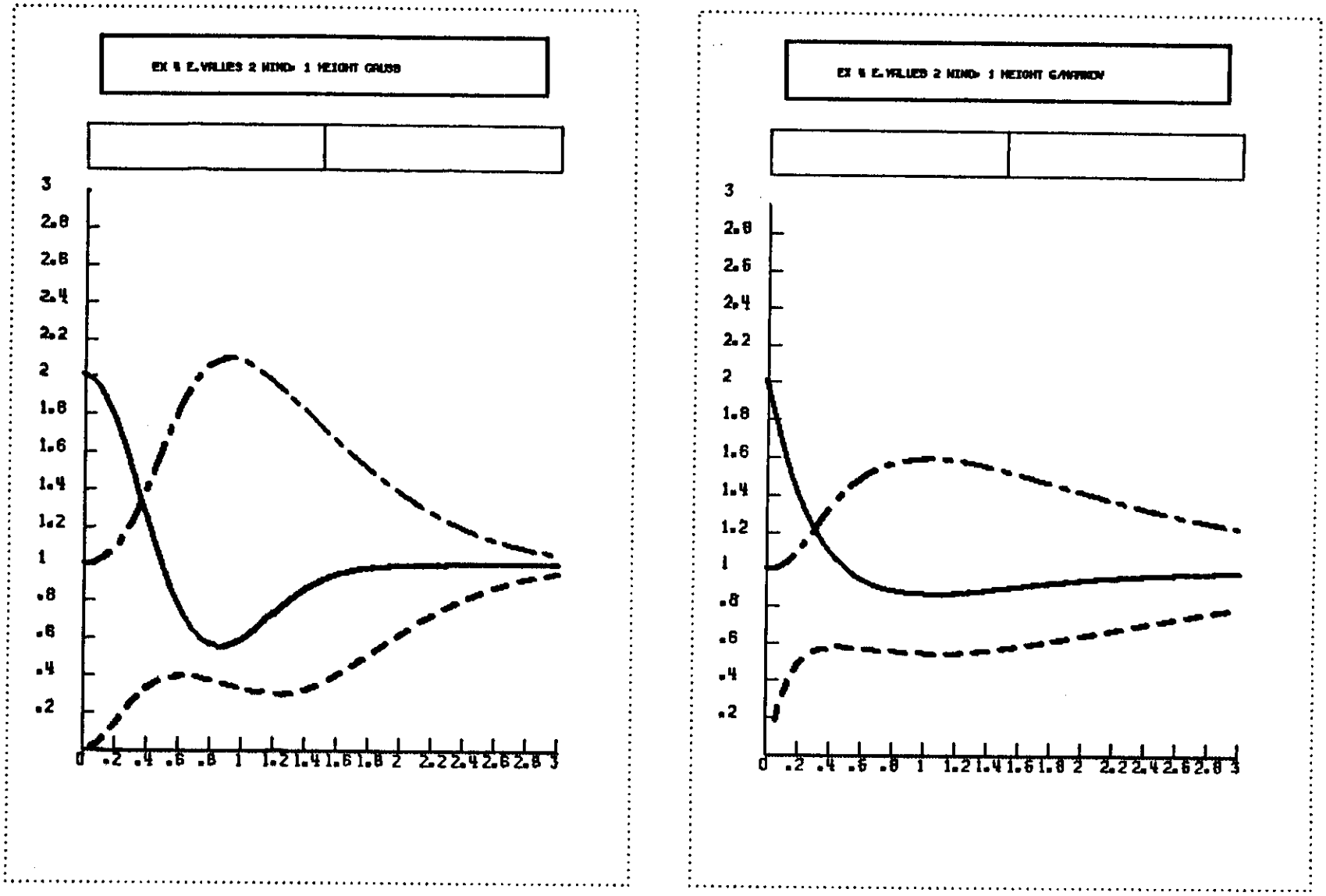


Figure 6.13 Plots of the dependence of the three eigenvalues of the fourth example (one height and two winds) on the normalised spatial separation parameter for the Gaussian (left) and the Gauss Markovian (right). The solid line is the mean-wind eigenvalue, the smallest eigenvalue is the anti-geostrophic one, while the third is the geostrophic eigenvalue.

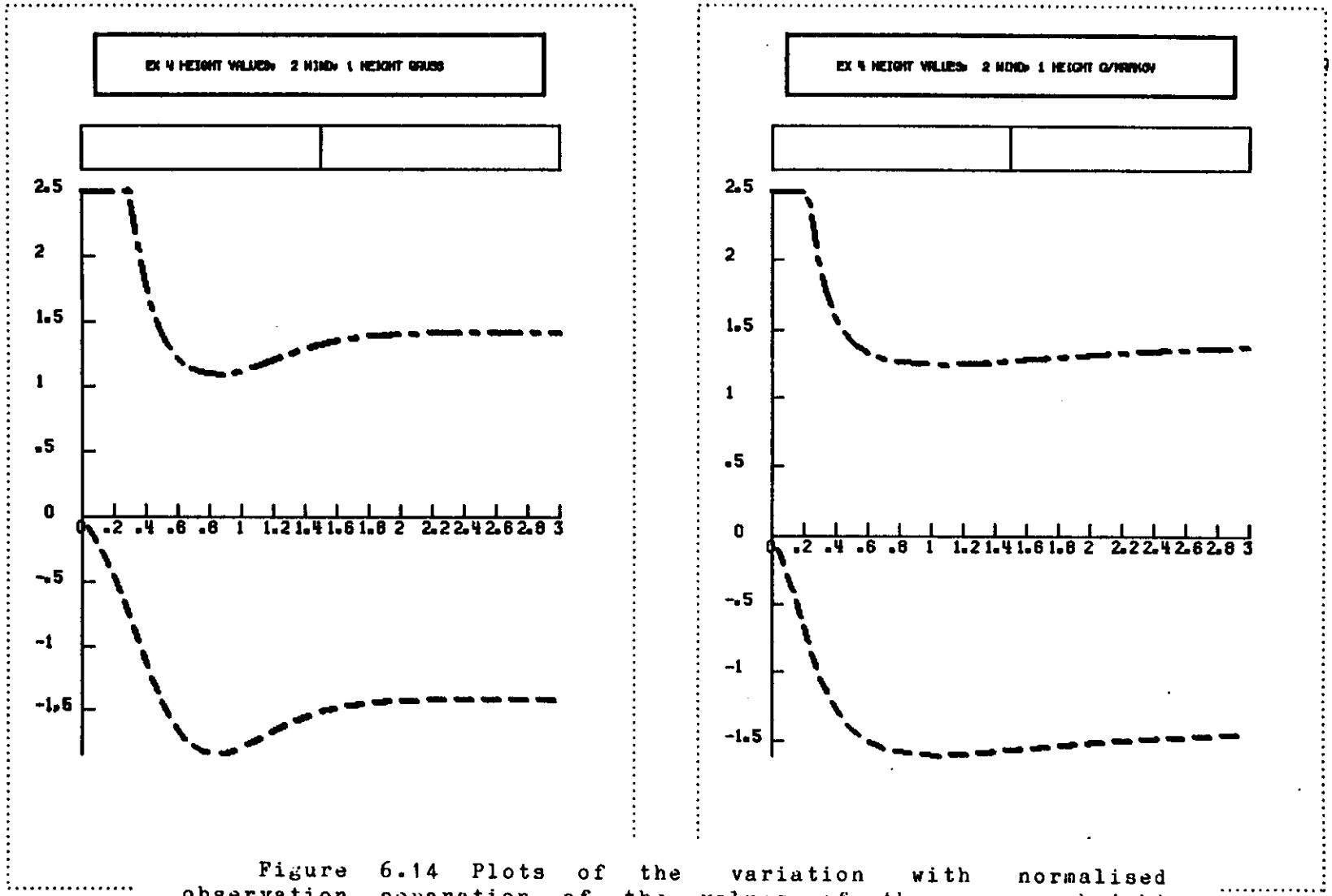


Figure 6.14 Plots of the variation with normalised observation separation of the values of the non-zero height elements in the eigenvectors of the fourth example (one height and two winds); results for the Gaussian on the left, Gauss-Markovian on the left. Geostrophic in this context means that positive relative vorticity in the eigenfunction is associated with low pressure at the height point between the winds.

INPUT P MATRIX

1	1.00	.61	.14	.01	.00	.00	.00	.00	.00
2	.61	1.00	.61	.14	.01	.00	.00	.00	.00
3	.14	.61	1.00	.61	.14	.01	.00	.00	.00
4	.01	.14	.61	1.00	.61	.14	.01	.00	.00
5	.00	.01	.14	.61	1.00	.61	.14	.01	.00
6	.00	.00	.01	.14	.61	1.00	.61	.14	.01
7	.00	.00	.00	.01	.14	.61	1.00	.61	.14
8	.00	.00	.00	.00	.01	.14	.61	1.00	.61
9	.00	.00	.00	.00	.00	.01	.14	.61	1.00

Table 6.1 The prediction error correlation matrix P for nine equally spaced collinear observations of height using a Gaussian structure function with non-dimensional observation spacing 1.

EIGENVALUES OF P

1	.55E-01
2	.12E+00
3	.25E+00
4	.47E+00
5	.78E+00
6	.12E+01
7	.17E+01
8	.21E+01
9	.24E+01

OB SPACING 1.00

Table 6.2 The eigenvalues of the matrix in table 6.1

EIGENVALUES OF P

9	.06	.12	.25	.47	.78	1.19	1.65	2.08	2.39

EIGENVECTORS OF P

	1	2	3	4	5	6	7	8	9
1	.11	.22	-.33	-.41	-.45	-.45	-.39	-.29	.16
2	-.25	-.42	.45	.32	.06	-.21	-.40	-.43	.27
3	.36	.44	-.18	.22	.44	.29	-.11	-.41	.36
4	-.44	-.28	-.25	-.43	-.02	.41	.27	-.25	.42
5	.46	.00	.45	.00	-.44	.00	.44	-.00	.44
6	-.44	.28	-.25	.43	-.02	-.41	.27	.25	.42
7	.36	-.44	-.18	-.22	.44	-.29	-.11	.41	.36
8	-.25	.42	.45	-.32	.06	.21	-.40	.43	.27
9	.11	-.22	-.33	.41	-.45	.45	-.39	.29	.16

Table 6.3 The eigenvectors and corresponding eigenvalues of the matrix in table 6.1

INVERTED MATRIX ASSUMING NO UNCOR NOISE

	1	2	3	4	5	6	7	8	9
1	1.98	-1.90	1.33	-.85	.52	-.31	.18	-.09	.04
2	-1.90	3.80	-3.18	2.14	-1.34	.81	-.47	.25	-.09
3	1.33	-3.18	4.70	-3.74	2.48	-1.53	.89	-.47	.18
4	-.85	2.14	-3.74	5.04	-3.93	2.56	-1.53	.81	-.31
5	.52	-1.34	2.48	-3.93	5.13	-3.93	2.48	-1.34	.52
6	-.31	.81	-1.53	2.56	-3.93	5.04	-3.74	2.14	-.85
7	.18	-.47	.89	-1.53	2.48	-3.74	4.70	-3.18	1.33
8	-.09	.25	-.47	.81	-1.34	2.14	-3.18	3.80	-1.90
9	.04	-.09	.18	-.31	.52	-.85	1.33	-1.90	1.98

Table 6.4 Inverse of the matrix in table 6.1

INVERTED MATRIX ALLOWING UNCOR NOISE

	1	2	3	4	5	6	7	8	9
1	1.08	-.62	.20	-.04	.00	.00	-.00	.00	-.00
2	-.62	1.43	-.73	.23	-.05	.00	.00	-.00	.00
3	.20	-.73	1.47	-.74	.23	-.04	.00	.00	-.00
4	-.04	.23	-.74	1.47	-.74	.23	-.04	.00	.00
5	.00	-.05	.23	-.74	1.47	-.74	.23	-.05	.00
6	.00	.00	-.04	.23	-.74	1.47	-.74	.23	-.04
7	-.00	.00	.00	-.04	.23	-.74	1.47	-.73	.20
8	.00	-.00	.00	.00	-.05	.23	-.73	1.43	-.62
9	-.00	.00	-.00	.00	.00	-.04	.20	-.62	1.08

Table 6.5 Inverse of the matrix $\underline{\underline{P}}+\underline{\underline{Q}}$ where $\underline{\underline{P}}$ is given in table 6.1 and $\underline{\underline{Q}}$ corresponds to equal amplitude uncorrelated errors with non-dimensional amplitude 0.5, variance 0.25.

INPUT P MATRIX

1	1.00	.88	.61	.32	.14	.04	.01	.00	.00
2	.88	1.00	.88	.61	.32	.14	.04	.01	.00
3	.61	.88	1.00	.88	.61	.32	.14	.04	.01
4	.32	.61	.88	1.00	.88	.61	.32	.14	.04
5	.14	.32	.61	.88	1.00	.88	.61	.32	.14
6	.04	.14	.32	.61	.88	1.00	.88	.61	.32
7	.01	.04	.14	.32	.61	.88	1.00	.88	.61
8	.00	.01	.04	.14	.32	.61	.88	1.00	.88
9	.00	.00	.01	.04	.14	.32	.61	.88	1.00

Table 6.6 The prediction error correlation matrix $\underset{\sim}{P}$ for nine equally spaced collinear observations of height using a Gaussian structure function with non-dimensional observation spacing 0.5

EIGENVALUES OF P

1	.26E-04
2	.45E-03
3	.43E-02
4	.27E-01
5	.13E+00
6	.47E+00
7	.13E+01
8	.27E+01
9	.43E+01

OB SPACING .50

Table 6.7 The eigenvalues of the matrix in table 6.6

EIGENVALUES OF P

9	.00	.00	.00	.03	.13	.47	1.31	2.75	4.31

EIGENVECTORS OF P

	1	2	3	4	5	6	7	8	9
1	.04	.10	.21	.33	-.44	.49	.47	-.37	-.20
2	-.15	-.33	-.47	-.46	.26	.06	.33	-.42	-.29
3	.32	.49	.35	-.05	.39	-.35	.03	-.37	-.36
4	-.48	-.38	.14	.42	-.09	-.37	-.28	-.22	-.40
5	.54	.00	-.43	-.00	-.41	.00	-.41	-.00	-.42
6	-.48	.38	.14	-.42	-.09	.37	-.28	.22	-.40
7	.32	-.49	.35	.05	.39	.35	.03	.37	-.36
8	-.15	.33	-.47	.46	.26	-.06	.33	.42	-.29
9	.04	-.10	.21	-.33	-.44	-.49	.47	.37	-.20

Table 6.8 The eigenvectors and corresponding eigenvalues of the matrix in table 6.6

INVERTED MATRIX ASSUMING NO UNCOR NOISE

	1	2	3	4	5	6	7	8	9
1	.87E+02	-.30E+03	.56E+03	-.72E+03	.72E+03	-.56E+03	.34E+03	-.14E+03	.32E+02
2	-.30E+03	.11E+04	-.22E+04	.29E+04	-.30E+04	.24E+04	-.15E+04	.62E+03	-.14E+03
3	.56E+03	-.22E+04	.44E+04	-.62E+04	.66E+04	-.54E+04	.34E+04	-.15E+04	.34E+03
4	-.72E+03	.29E+04	-.62E+04	.91E+04	-.10E+05	.85E+04	-.54E+04	.24E+04	-.56E+03
5	.72E+03	-.30E+04	.66E+04	-.10E+05	.11E+05	-.10E+05	.66E+04	-.30E+04	.72E+03
6	-.56E+03	.24E+04	-.54E+04	.85E+04	-.10E+05	.91E+04	-.62E+04	.29E+04	-.72E+03
7	.34E+03	-.15E+04	.34E+04	-.54E+04	.66E+04	-.62E+04	.44E+04	-.22E+04	.56E+03
8	-.14E+03	.62E+03	-.15E+04	.24E+04	-.30E+04	.29E+04	-.22E+04	.11E+04	-.30E+03
9	.32E+02	-.14E+03	.34E+03	-.56E+03	.72E+03	-.72E+03	.56E+03	-.30E+03	.87E+02

Table 6.9 Inverse of the matrix in table 6.6

INVERTED MATRIX ALLOWING UNCOR NOISE

	1	2	3	4	5	6	7	8	9
1	1.64	-1.17	-.14	.21	.09	-.04	-.04	.00	.01
2	-1.17	2.48	-1.07	-.29	.14	.12	-.01	-.04	.00
3	-.14	-1.07	2.49	-1.09	-.30	.14	.12	-.01	-.04
4	.21	-.29	-1.09	2.51	-1.08	-.30	.14	.12	-.04
5	.09	.14	-.30	-1.08	2.52	-1.08	-.30	.14	.09
6	-.04	.12	.14	-.30	-1.08	2.51	-1.09	-.29	.21
7	-.04	-.01	.12	.14	-.30	-1.09	2.49	-1.07	-.14
8	.00	-.04	-.01	.12	.14	-.29	-1.07	2.48	-1.17
9	.01	.00	-.04	-.04	.09	.21	-.14	-1.17	1.64

Table 6.10 Inverse of the matrix $\underset{\sim}{P}+\underset{\sim}{Q}$ where $\underset{\sim}{P}$ is given in table 6.6 and $\underset{\sim}{Q}$ corresponds to equal amplitude uncorrelated errors with non-dimensional amplitude 0.5, variance 0.25.

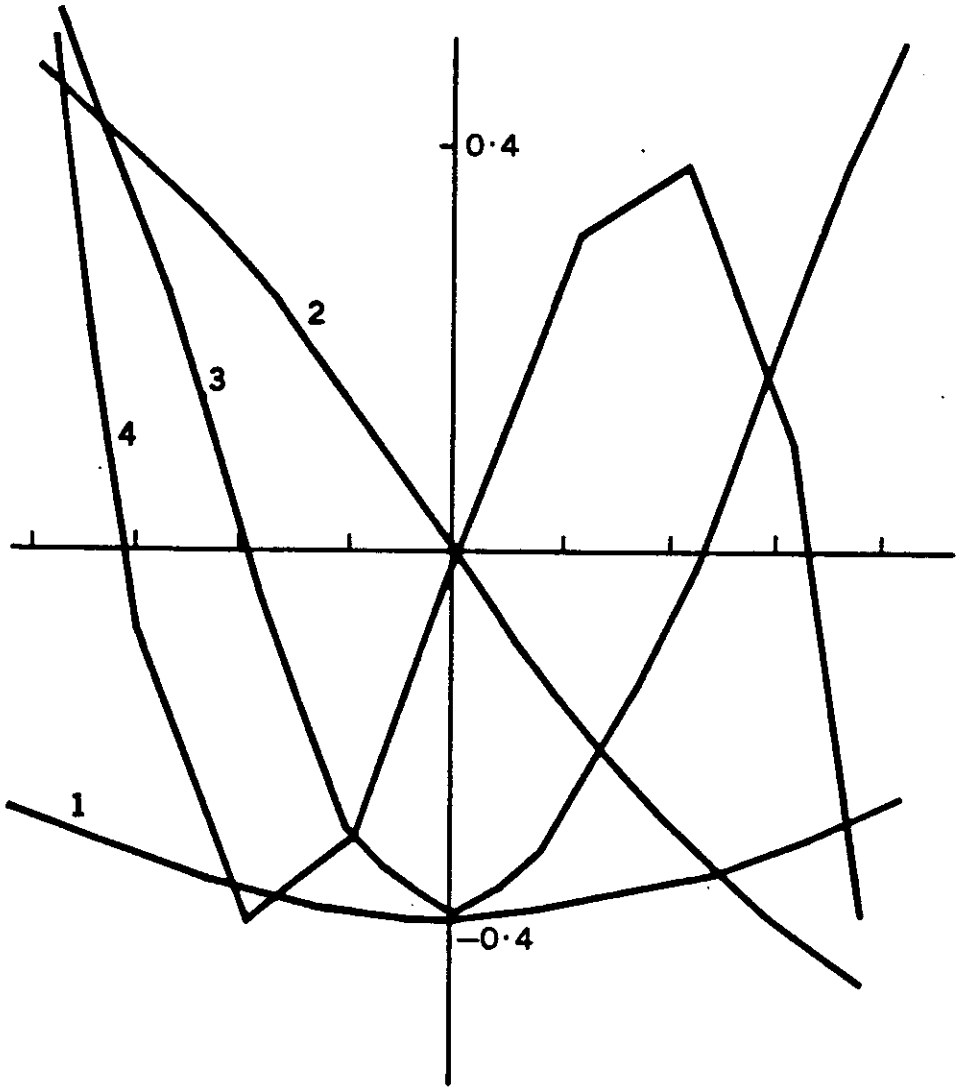


Figure 6.15 The first four eigenvectors of the matrix for the problem of the analysis of height at nine collinear points using the Gaussian as the structure function.

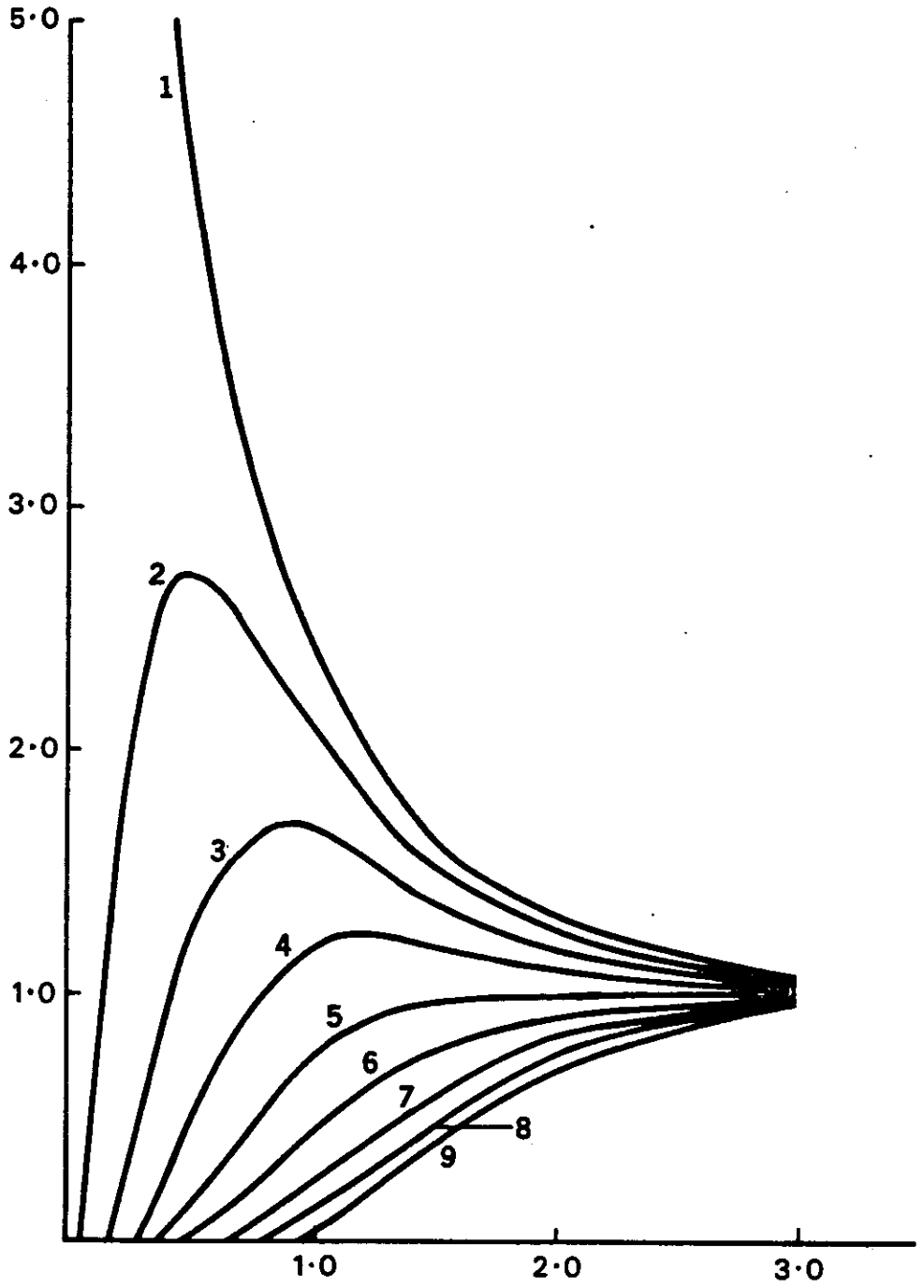


Figure 6.16 Plot of the variation with normalised observation separation of the eigenvalues of the same problem as fig 6.15

6.3 Numerical examples. O.I. as a smoothing/desmoothing operator

In the last Section we saw some examples of how the type of variable to be analysed, and the ratio of observation spacing to structure function width, leads to considerable variations in the resolution of the O.I. system as measured by the magnitude of the eigenvalue for particular eigenmodes. In the present section we shall verify some of these results in slightly more complex univariate cases. We also examine the internal structure of the O.I. matrix where we find some suggestions that the procedure has an interpretation in terms of smoothing/desmoothing operators.

6.3.1 Univariate analysis with a Gaussian correlation. Numerical example

Let us consider the case of equally spaced observations where we model the correlation function by a Gaussian $P_{ij} = \exp -\frac{1}{2}\left(\frac{(i-j)\Delta x}{b}\right)^2$.

This is what is used operationally, if only height observations are available and we only wish to do a height analysis.

This structure function depends on the ratio of two parameters Δx, the observation spacing and b, the width of the Gaussian. Table 6.1 shows the $\underline{P}$ matrix in the case of nine equally spaced observations with $\Delta x/b = 1$. The eigenvalues are shown in Table 6.2, the eigenvectors in Table 6.3 the inverse matrix $(\underline{P}^{-1})$ in Table 6.4 and $(\underline{P}+\underline{Q})^{-1}$ in Table 6.5 when $\underline{Q} = \sigma^2 \underline{I}$, $\sigma^2 = 0.25$.

Fig.6.15 shows a plot of the four gravest eigenvectors of $\underline{P}$. The gravest eigenvector (the one corresponding to the largest eigenvalue) has no zeros. In the limit $N \to \infty$ this component of the data would be related to the mean value of the data. In such a case the best analysed part of the data is the mean value. The second eigenvector has one zero and looks like a linear trend in the data. The response of the analysis to such a feature in the data is also rather good. We continue in this way to the highest eigenvector, which looks like a two grid wave. If this component occurs in the data it will be heavily damped.

6.3.2 Effect of observation spacing or width of structure function

In order to examine the effect of observation spacing (or equivalently, the width of the structure function, since it is the ratio $\Delta x/b$ which matters) we present in Tables 6 to 10 the results corresponding to Tables 6.1 to 6.5 except that in this case the observation spacing is changed so that $\Delta x/b = 0.5$, rather than 1 as in the previous case. The change has little impact on the eigenvectors as they still have essentially the same structures. However the effect on the eigenvalues is quite dramatic. The following calculations are done with σ^2, the observational error (normalised) set at $\sigma^2 = 0.25$.

Table 6.11. The ratio $\frac{\lambda}{\lambda+\sigma^2}$ for the cases $\frac{\Delta x}{b} = 1.0,\ 0.5$

Mode	$\Delta x/b=1.0$	$\Delta x/b=0.5$
1	.215	.00001
2	.302	.0018
3	.5	.017
4	.65	.10
5	.76	.34
6	.83	.65
7	.87	.84
8	.89	.92
9	.91	.95

Table 6.11 shows the ratios $\frac{\lambda}{\lambda+\sigma^2}$ for the two situations. When we halve the observation spacing, or equivalently, double the width of the structure function, we make radical changes to the ability of the analysis to draw for grid scale structure in the data. Table 6.11 provides a measure of this ability. Fig 6.16 plots the values of the eigenvalues (for $0<\Delta x/b<3$) for this nine point problem. The eigenvectors do not change character at any point. If the data is closely spaced then the "mean-value" eigenvector dominates. When $\Delta x/b \sim 1$ then half the eigenvalues are larger than 1 and the other half less than 1. For large separation all the eigenvalues asymptote to 1.

A slightly different view of these results can be had by considering the analysis error calculation

$$\varepsilon_k^{12} = 1 - (\underline{M}^{-1}\ \underline{P}_k)\ \underline{P}_k \text{ of Equation 5.2.8.}$$

It is an easy matter to show that this result can be generalised to provide the analysis error covariance matrix

$$\underline{E}_a = \underline{E}[1 - \frac{\lambda}{\lambda+\sigma^2}]\ \underline{E}^T$$

$$= \underline{E}[\frac{\sigma^2}{\sigma^2+\lambda}]\ \underline{E}^T$$

This says that the analysis error in the well resolved scales is very small, since $\sigma^2 << \lambda$, while the error in the poorly resolved scales is essentially the same as the first guess error in these scales, i.e. 1.

There is a clear suggestion in these results, which is verified in practice, that the O.I. analysis will have quite definite limitations as to resolution, once the parameters of the problems, the first guess error, observation error, observation spacing and scale of the structure function have been fixed. Structures in the data on small scales will be treated as noise and smoothed. On the other hand we note that if the observational error is zero then the analysis will draw exactly for the data. This is, of course what one should do for exact data.

In addition to its other advantages, the ability of the OI formalism to give a rigorous account of the analysis error and the resolution is a considerable help in the study of ways of improving the analysis.

6.3.3 A finite difference interpretation of the O.I. results

Any given row of the matrices $\underline{P}$ in Table 6.1 could be thought of as a finite difference operator corresponding to an amplifying integral of the form $\sum_1^N w_i B_i$ where the weights w_i are given by the entries in the matrix $\underline{P}$. Correspondingly the matrix $(\underline{P}+\underline{Q})^{-1}$ in Table 6.5 has a suggestive structure; it appears rather like a damped finite difference operator $\frac{1}{4}(-1, 2.5, -1)$. We

should note however that $\underline{P}^{-1}$ is rather far from having this simple structure (Table 6.4). When we allow for observational error the O.I. operator appears to be the product of an averaging operator $\underline{P}$ and a de-smoothing operator $(\underline{P}+\underline{Q})^{-1}$. The eigenvalues of $\underline{P}^T\ \underline{M}^{-1}$ are ordered in the same way as those of $\underline{P}$.

The largest response to an integrating operator is given by a slowly varying (large scale, nearly constant) term while a rapidly oscillating term will give a much smaller response. Thus on qualitative grounds we should expect that if $\underline{P}$ has the form of an averaging operator then the O.I. operator should give the largest response to the slowly varying data while it gives the smallest response to rapidly varying data. We shall see in what follows that the converse is also true. If $\underline{P}$ is of the form of a derivative operator then the O.I. matrix gives the largest response to rapidly varying data and the smallest response to slowly varying data.

6.3.4 Numerical example: univariate wind analysis

We consider the problem of analysing 9 collinear wind observations and suppose that the observed wind is normal to the line of observation points.

The structure function that we will use is

$$\overset{\approx}{g}(\xi) = (1-\xi^2)\ e^{-1/2(\xi^2)}$$

where g is the Gaussian. Let us suppose the observation spacing is such that $\frac{\Delta x}{b} = \xi = \sqrt{3}$, so that a given observation is in the negative lobe of its neighbours' structure function. The $\underline{P}$ matrix is shown in Table 6.12, and clearly looks like a (-1, 2.2, -1) finite difference operator, at least in the centre of the range. The eigenvalues and eigenvectors are shown in Table 6.14. The largest eigenvalue clearly corresponds to the two grid wave while the smallest eigenvalue corresponds to the "mean-value" eigenvector. This is just what we found in our earlier three-point examples.

INPUT P MATRIX

1	1.00	-.45	-.03	-.00	-.00	-.00	-.00	-.00	-.00
2	-.45	1.00	-.45	-.03	-.00	-.00	-.00	-.00	-.00
3	-.03	-.45	1.00	-.45	-.03	-.00	-.00	-.00	-.00
4	-.00	-.03	-.45	1.00	-.45	-.03	-.00	-.00	-.00
5	-.00	-.00	-.03	-.45	1.00	-.45	-.03	-.00	-.00
6	-.00	-.00	-.00	-.03	-.45	1.00	-.45	-.03	-.00
7	-.00	-.00	-.00	-.00	-.03	-.45	1.00	-.45	-.03
8	-.00	-.00	-.00	-.00	-.00	-.03	-.45	1.00	-.45
9	-.00	-.00	-.00	-.00	-.00	-.00	-.03	-.45	1.00

Table 6.12 The prediction error correlation matrix $\underset{\approx}{P}$ for nine equally spaced collinear observations of the normal component of the wind using a Gaussian based non-divergent structure function with non-dimensional observation spacing 1.73

EIGENVALUES OF P

1	.11E+00
2	.26E+00
3	.48E+00
4	.76E+00
5	.10E+01
6	.13E+01
7	.15E+01
8	.17E+01
9	.18E+01

OB SPACING 1.73

Table 6.13 The eigenvalues of the matrix in table 6.12

EIGENVALUES OF P

9	.11	.26	.48	.76	1.04	1.31	1.53	1.70	1.80

EIGENVECTORS OF P

	1	2	3	4	5	6	7	8	9
1	.14	-.27	-.37	-.43	-.45	.42	-.35	-.25	.13
2	.26	-.43	-.42	-.25	.02	-.28	.43	.42	-.26
3	.36	-.42	-.13	.27	.45	-.25	-.15	-.43	.36
4	.42	-.26	.27	.42	-.01	.43	-.26	.27	-.43
5	.44	-.00	.45	-.00	-.45	.00	.45	.00	.45
6	.42	.26	.27	-.42	-.01	-.43	-.26	-.27	-.43
7	.36	.42	-.13	-.27	.45	.25	-.15	.43	.36
8	.26	.43	-.42	.25	.02	.28	.43	-.42	-.26
9	.14	.27	-.37	.43	-.45	-.42	-.35	.25	.13

Table 6.14 The eigenvectors and corresponding eigenvalues of the matrix in table 6.12

INVERTED MATRIX ASSUMING NO UNCOR NOISE

	1	2	3	4	5	6	7	8	9
1	1.46	.99	.73	.53	.38	.27	.18	.11	.06
2	.99	2.14	1.49	1.09	.78	.55	.38	.23	.11
3	.73	1.49	2.49	1.74	1.26	.89	.60	.38	.18
4	.53	1.09	1.74	2.66	1.84	1.31	.89	.55	.27
5	.38	.78	1.26	1.84	2.71	1.84	1.26	.78	.38
6	.27	.55	.89	1.31	1.84	2.66	1.74	1.09	.53
7	.18	.38	.60	.89	1.26	1.74	2.49	1.49	.73
8	.11	.23	.38	.55	.78	1.09	1.49	2.14	.99
9	.06	.11	.18	.27	.38	.53	.73	.99	1.46

Table 6.15 Inverse of the matrix in table 6.12

INVERTED MATRIX ALLOWING UNCOR NOISE

	1	2	3	4	5	6	7	8	9
1	.95	.42	.21	.10	.05	.02	.01	.01	.00
2	.42	1.13	.51	.25	.12	.06	.03	.01	.01
3	.21	.51	1.18	.53	.26	.13	.06	.03	.01
4	.10	.25	.53	1.19	.53	.26	.13	.06	.02
5	.05	.12	.26	.53	1.19	.53	.26	.12	.05
6	.02	.06	.13	.26	.53	1.19	.53	.25	.10
7	.01	.03	.06	.13	.26	.53	1.18	.51	.21
8	.01	.01	.03	.06	.12	.25	.51	1.13	.42
9	.00	.01	.01	.02	.05	.10	.21	.42	.95

Table 6.16 Inverse of the matrix $\underset{\sim}{P}+\underset{\sim}{Q}$ where $\underset{\sim}{P}$ is given in table 6.12 and $\underset{\sim}{Q}$ corresponds to equal amplitude uncorrelated errors with non-dimensional amplitude 0.5, variance 0.25.

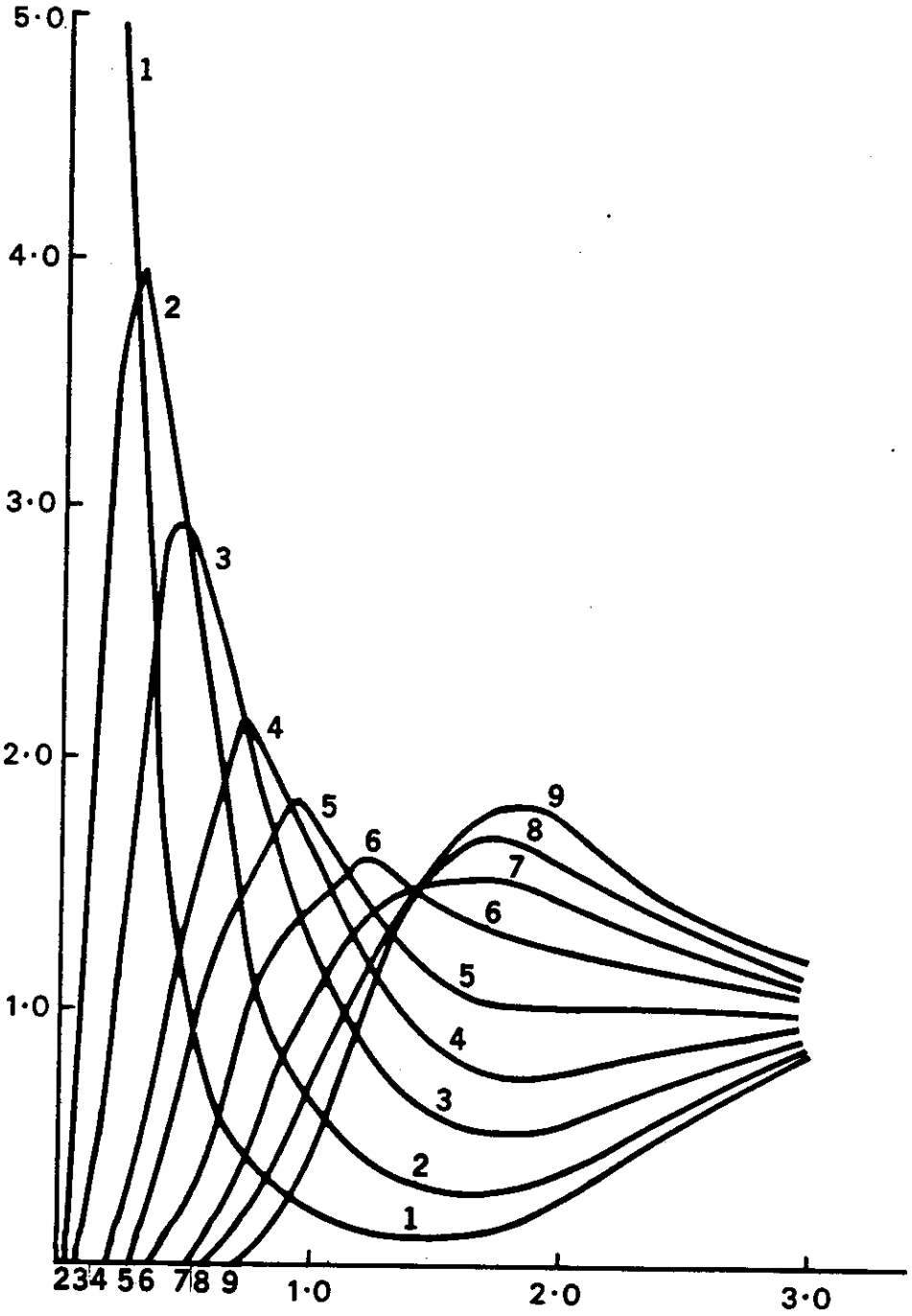


Figure 6.17 Plot of the variation with normalised observation separation of the eigenvalues of the problem of analysing the normal wind along a line of nine points using Gaussian-based non-divergent structure functions.

Fig.6.17 shows the behaviour of the eigenvalues of this problem. The eigenvalues are labelled by their order at very small separation. The discontinuities we saw in the earlier, simple, example are not evident but the general behaviour is very similar. As the observation separation increases to $\Delta x/b \sim 1$ the ordering of the eigenvalues changes radically so that by the time $\xi = \sqrt{3}$, the ordering at small separation is completely reversed, with the two grid wave having the largest eigenvalue.

To complete this Section we examine briefly the structure of the inverse matrix $\underline{\underline{N}}^{-1}$. We have already seen that the $\underline{\underline{P}}$ matrix looks like a second order finite difference operator, because we've chosen the separation of the observations to be such that each observation lies at the minimum of its neighbours' negative lobes. The matrix $\underline{\underline{P}}^{-1}$ is shown in Table 6.15 and it has the character of an averaging operator. Table 6.16 contains the inverse matrix $\underline{\underline{N}}^{-1}$, which also has a similar character. It is intuitively clear then that for separations >1 for which all the entries in $\underline{\underline{P}}$ will be negative, except for the main diagonal, the largest eigenvalues of $\underline{\underline{P}}$ will correspond to rapidly varying structure.

6.4 Discussion

The results presented above contain very simple examples, but they serve to illustrate a number of important points.

The basic result is Eq.6.11.

$$\beta_k^i = \beta_k \frac{\lambda_k}{\lambda_k + \sigma^2}$$

which shows that some components of the data are returned undamped by the analysis while others are severely damped. In a typical operational application, with say 100 data, then perhaps 10% of the eigenvectors will be larger than 1, while σ^2 will have values of the order of .25 (Lönnberg 1983). Typically the large eigenvalues will correspond to the larger scales in the data. Then we would expect to find about 10% of the modes in the data to be

essentially undamped, about 50% of the modes to be significantly damped, since they are treated as noise, while the remaining 40% will be damped to some degree, depending on the values of λ_i and σ_2.

The distribution of the eigenvalues is very sensitive to the variables which were observed, and to the constraints placed on the analysis. We have seen in our simple calculations some examples which are of quite general validity. Firstly we see that the constraint of geostrophy leads to a damping of any anti-geostrophic components of the data. Such components will always occur because of instrumental error. It is important to understand the difference between geostrophy in the sense in which we use it here, and the subtly different way in which it is used by Lorenc (1981). Lorenc says that the OI system produces strictly geostrophic increments if we impose a strictly geostrophic constraint. This statement is true for the continuous functional representation underlying the analysis. However, if we discretize this continuous analysis on a set of observation points then there will be components in the discretised analysis which are anti-geostrophic in terms of finite differences on the observation grid. For this reason, too, it is important for the efficacy of the constraints that the grid on which the analysis is evaluated should have several grid-lengths per scale length of the structure function.

The discussion of the wind analysis led to a counter-intuitive result, namely that if the data is too widely spaced, then the larger scales are badly under-analysed. This result stemmed from the constraint of non-divergence. The general significance of this result is still not fully understood. It may be that it does summarise some unpleasant problems in the large-scale tropical analyses. An extensive discussion of the constraints in an OI analysis may be had in the monograph by Daley (1983).

We make a few comments about the structure functions themselves. It is clear that all the properties of the OI matrix are determined by the structure functions for prediction and observational error. If the analysis is to be optimal then we have to determine these quantities with some care. Roughly speaking, the observations tell us about forecast errors, and the structure functions tell us how to spread out the information in the observations in an optimal way. If the functions we use are too broad compared to what is justified, then we fail to resolve real features in the data. A similar effect will occur if the observational errors are set too high relative to the forecast errors, as can be seen from 6.11. From the same arguments, if the structure functions are too narrow, or the observational errors are set too low then we will not do enough smoothing, and we will draw for observational error.

The effect of the constraints is very important for large scale analysis. If we treat the height data and the two components of the wind data as uncorrelated, then we will give equal weight to the geostrophic and a-geostrophic components, to the divergent and non-divergent components of the data. Such an analysis, if used to start a forecast, will excite extensive gravity wave activity which will rapidly disperse much of the information we have tried to convey to the model. The filtering properties of the OI analysis are very important as they allow one to control the degree of geostrophy or non-divergence in the analysis increments as desired. For example there is some evidence that we ought to permit a small degree of divergence into the increments; the OI formalism offers an elegant way of doing just that.

Finally one must qualify everything above by saying that all the comments apply to the analysis increments, and not the analysis, which is the sum of the increments and the first guess. The atmosphere does not satisfy simple linear laws such as non-divergence of the wind, or geostrophic balance. Non-

linear laws such as gradient-wind laws are often much more accurate descriptions of the atmosphere. However if one has a very accurate first guess, then the OI procedure needs only deal with small perturbations to a non-linearly balanced state. These perturbations can then be treated, with good accuracy, as satisfying linear laws. Thus the accuracy of the first-guess is an important prerequisite for the validity of the linear constraints which are used in OI. Currently we are approaching the stage where the first-guess errors and the observation errors are of similar magnitude. This has many interesting consequences, not least in the area of long term quality control of the data itself.

References

Bergman, K.H. 1979: Multivariate analysis of temperatures and winds using optimum interpolation. Mon.Wea.Rev., 107, 1423-1444.

Bergthorsson, P. and Döös, B.R. 1955: Numerical weather map analysis, Tellus, 7, 329-340.

Cressman, G.P., 1959: An operational objective analysis system. Mon.Wea.Rev., 87, 367-374.

Daley, R., 1983: The spectral characteristics of the ECMWF objective analysis system. ECMWF Tech.Rep. in press.

Döös, B.R. 1969: Numerical analysis of meteorological data. Lectures on numerical short-range weather prediction, Moscow, 1965, pp.679-706. Pub. Hydrometeoizdat, Leningrad 1969.

ECMWF Workshop on the use of empirical orthogonal functions in meteorology, 1977.

Flattery, T.W., 1971: Spectral models for global analysis and forecasting. Proc.Sixth AWS Technical Exchange Cont.U.S. Naval Academy, 1970, Air Weather Service. Tech.Report, 142 pp. 42-53.

Franke, R. and Gordon, W.J., 1983: The structure of optimum interpolation functions. Naval Postgraduate School, Monterey, Calif. NPS-53-83-005 (Feb.83).

Gandin, L.S., 1963: Objective analysis of meteorological fields. Translated from Russian by the Israeli Program for Scientific Translations (1965).

Gilchrist, B. and Cressman, G.P., 1954: An experiment in objective analysis. Tellus, 6, 309-318.

Julian, P.R. and Thiebaux, H.J., 1975: On some properties of correlation functions used in optimum interpolation schemes. Mon.Wea.Rev., 103, 605-616.

Kasahara, A., 1976: Normal modes of ultralong waves in the atmosphere. Mon.Wea.Rev., 104, 669-690.

Lorenc, A.C. 1981: A global three-dimensional multivariate statistical interpolation scheme. Mon.Wea.Rev., 109, 701-721.

Lönnberg, P., 1983: Structure functions and their implications for higher resolution analysis. ECMWF Workshop on Current Problems in Data Assimilation.

Sasaki, Y., 1958: An objective analysis based on the variational method. J.Met.Soc.Japan, 36, pp.77-88.

Sasaki, Y., 1970: Some basic formalisms on numerical variational analysis. Mon.Wea.Rev., 98, 875-883.

Seaman, R.S., 1977: Absolute and differential accuracy of analyses achievable with specified observational network characteristics. Mon.Wea.Rev., 105, 1211-1222.

Temperton, C., 1973: Some experiments in dynamic initialization for a simple primitive equation model. Quart.J.Roy.Met.Soc., 99, 303-319.