287

MET O 11 TECHNICAL NOTE NO 244

Optimal Non-linear Objective Analysis

by

Andrew C Lorenc

Met O 11(Forecasting Research)
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SZ
ENGLAND.

November 1986.

FH2A

Optimal Non-linear Objective Analysis.

By Andrew C Lorenc

Meteorological Office, Bracknell.

November 1986

SUMMARY

Some practical problems in objective analysis for numerical weather prediction are best solved by using non-linear analysis equations. They include the utilization of non-linear prior constraints on the analysis, and the use of observations which are non-linearly related to the analysis parameters, or which have non-Gaussian error distributions. Bayesian methods are used to derive equations for the optimal (maximum likelihood) non-linear analysis. It is shown how to incorporate a strong constraint that the four-dimensional evolution of the analysis should be consistent with a NWP model, by reduction of the control variable to the space-dimensioned initial field for the model. The iterative solution of the non-linear analysis equation then involves the integration of the NWP model, and its adjoint.

The behavior of the non-linear equations is demonstrated with a simple one-dimensional shallow-water model. It is shown that time-tendency information, and indirect observations such as wind speed, or the movement of a tracer, can be used in the analysis. The resulting forecasts are better than those made from an analysis from a traditional analysis-forecast cycle. The non-linear method is shown to be capable of "moving" a discontinuity similar to a front, to fit observations defining its position, thus giving an analysis with more detail than would be expected from the spatial resolution of the

observations. The incorporation of additional non-linear constraints, such as that used in initialization, is demonstrated. The method can be used to effectively reject observations with gross errors, by specifying a non-Gaussian error distribution. However this generates multiple minima which complicate the search for the best analysis, so the complex decision taking algorithms associated with other methods of quality control are not avoided.

The convergence properties of iterative methods of solution, and approximations to the ideal equations, are studied, in order to provide some indication as to whether the non-linear effects might be allowed for in a practical analysis scheme.

1. INTRODUCTION

Most methods of objective analysis in use for meteorological data are linear in the observed data values. That is, each analysed value xa[j] can be written as a linear combination of the data values yo[i]; this is usually done in terms of deviations from background values xb[j] and yb[i]:

$$xa[j]=xb[j]+\sum_i W[j,i]\ (yo[i]-yb[i]) \qquad (1)$$

(1) is linear if the weights W[j,i] are independent of the actual observed values yo, but depend only on their positions and accuracies. This is true of least-squares fitting techniques (for instance fitting splines), of the successive-correction method, and of optimal interpolation. These methods are all related, and can be shown to be equivalent to Bayesian estimation, assuming that error distributions of both observations and of any background fields are Gaussian, prior constraints and relationships describing the desired analysis are linear, and the relationships between analysed and observed variables are linear (Lorenc 1986). Such assumptions lead to a variational problem with a quadratic (L2) penalty function. However this is not true of many of the problems currently of interest in meteorological objective analysis for numerical weather prediction (NWP). These problems include :-

a. Four-dimensional analysis constrained by non-linear prognostic equations, to use time-tendency information from observations.

b. Use of "indirect" observations, with a non-linear relationship between observed parameters and those analysed.

c. Analysis of near-discontinuities, such as fronts, which can be

regarded as a strong non-linear coupling between spatial scales in the atmosphere.

d. Multivariate analysis, subject to non-linear balance constraints.

e. Quality control of observations, and use of data whose error distributions are known to be non-Gaussian.

These problems have usually been tackled either by modification of the optimal linear equation, or by new techniques which often do not give optimal weighting to background information. Examples, classified as above, are: (a) Lewis and Derber (1985), Talagrand and Courtier (1986), Hoffmann (1986), (b) much work on temperature retrievals from satellite radiances, and Julian (1984), (d) Williamson and Daley (1983), Wahba (1982), (e) Purser (1984), Lorenc and Hammon (1986).

In section 2 of this paper we derive the equation for a non-linear optimal (maximum-likelihood Bayesian) analysis, following Lorenc (1986). This equation can in principle be solved iteratively, although practical meteorological examples are usually so large that this is not feasible. In order to study the full non-linear optimal equation, we apply it to a simple one-dimensional non-linear shallow-water equation model, described in section 3. The initial conditions used are such that a hydraulic jump develops; this is about the simplest system that simulates some of the features of atmospheric fronts, and with it we can study all the problems (a) to (e) listed above. This is done in section 4, referring each example to analagous practical problems. It is shown that all of the examples can be solved using the optimal non-linear analysis equation, so that in

principle at least the practical problems are capable of solution. Even for the simple example system, there are practical problems in finding the optimal solution. These are illustrated in section 5 for various descent algorithms, and implications of some approximations which might be made to the analysis equations for a practical problem are discussed. Finally in section 6 we summarize.

2. OPTIMAL ANALYSIS EQUATION

2a. Bayesian derivation.

Lorenc (1986) discusses optimal analysis for numerical weather prediction, and uses a Bayesian argument to derive equations to be satisfied by the most likely analysis. The reader is referred to that paper for further discussion of the derivation, which is repeated here.

We define the best analysis as the most likely vector xa in a space of possible representations of the atmosphere. These representations have their resolution determined by the requirements of the forecast model used for numerical weather prediction. The analysis equations which we shall derive are general, for example they might be used for spectral or grid-point representations in two- three- or four-dimensions; in this paper we shall discuss them applied to space- and time-dimensioned grid-point represention s. Elements of this analysis space are all denoted by x, with a subscript as appropriate. Observations are not necessarily simply related to the elements of x. The available observations define a space of vectors y, with the

actual observed values defining vector yo. We assume that the observation and analysis spaces are precisely defined, so that knowledge of the true atmospheric state defines true analysis and observation vectors xt and yt. For example the precise meaning of a grid-point value as a space-average, and the space- and time-averaging properties of the observation instruments must be known. We further assume that the analysis space's representation of the atmosphere is sufficiently complete that given any x we can compute the most likely corresponding y using:

$$y = Kn(x) \tag{2}$$

The inverse of this, computing the most likely xa given the observed values yo, is our analysis problem. Because the observations have observational errors and are insufficient to determine the best x, we need to use prior information about the probability that any state x is the true state. This prior information contains a prior estimate of the most likely true x, the background state xb. Since in most of the following we shall assume that the prior probability density that x is the true state is a function of x-xb, we shall denote that the prior probability density function by Pb(x-xb). Because of instrumental errors, yo does not define yt; we denote the probability density function that y is the true state by Po(y-yo). We must also allow for the fact that neither the analysis space nor the observation space are full representations of the true atmosphere, so that in general Kn will have errors. Thus knowing xt only defines a probability density function for yt, which we denote by Pf(y-Kn(xt)). We shall need to know the probability of getting a particular set of

observed values yo, given a particular true analysis state xt. This requires the convolution of Pf with Po. The resulting probability distribution function, which we denote by Pof, describes both the errors of representativeness and the instrumental errors; often the combination is referred to simply as the observational error. Pof is given by:

$$Pof(Kn(xt)-yo) = \int Po(yo-y)\ Pf(y-Kn(xt))\ dy \qquad (3)$$

We are now in a position to use Bayes' Theorem to calculate the probability Pa(x) that any state x is the true state, given a particular set of observed values yo, and the background information described by Pb(x-xb). Pa(x) is the analysis probability distribution, given by:

$$Pa(x) = Pof(Kn(x)-yo)\ Pb(x-xb) \Big/ \int Pof(Kn(x1)-yo)\ Pb(x1-xb)\ dx1 \qquad (4)$$

The maximum likelihood estimate of the "best" analysis xa is then defined as that x which maximizes Pa(x). The product of probabilities which appears in (4) can be transformed into a sum by taking logarithms. The normalization term in the denominator is constant, and does not affect xa. Thus maximizing (4) is equivalent to minimizing a penalty function J given by:

$$J(x) = -\ln(Pof(Kn(x)-yo)) - \ln(Pb(x-xb)) \qquad (5)$$

2b. Optimum linear analysis

It is instructive at this stage to relate (5) to the better known optimal interpolation (OI) equations. To derive these using (5) we need a linearization of Kn whch is valid in the region of xb, and require the probability distribution functions to be Gaussian. Assuming that Kn is differentiable, with Jacobian matrix K, we get:

$$Kn(x+dx) = Kn(x) + K\,dx \tag{6}$$

Multidimensional Gaussian distributions are:

$$Pf(y-Kn(x)) = \exp(-(y-Kn(x))^{*}\,F^{-1}\,(y-Kn(x))/2) \tag{7}$$

$$Po(y-yo) = \exp(-(y-yo)^{*}\,O^{-1}\,(y-yo)/2) \tag{8}$$

$$Pb(x-xb) = \exp(-(x-xb)^{*}\,B^{-1}\,(x-xb)/2) \tag{9}$$

The superscript * denotes the Hermitian conjugate, or adjoint; for our real vectors and matrices the transpose. F O and B are the covariance matrices for errors of representativeness, instrumental errors, and background errors respectively. (The normalization constants to ensure that these probability distribution functions integrate to unity have been omitted since they do not affect the minimization). Substitution of (7) and (8) in (3) gives, to within a multiplicative constant:

$$Pof(Kn(x)-yo) = \exp(-(Kn(x)-yo)^{*}\,(O+F)^{-1}\,(Kn(x)-yo)/2) \tag{10}$$

Substitution of (9) and (10) in (5) gives:

$$J(x) = ((Kn(x)-yo)^{*} (O+F)^{-1} (Kn(x)-yo) + (x-xb)^{*} B^{-1} (x-xb))/2 \quad (11)$$

We denote partial derivatives with respect to the elements of the argument of a function by a prime '. Then:

$$J'(x) = K^{*} (O+F)^{-1} (Kn(x)-yo) + B^{-1} (x-xb) \quad (12)$$

$$J''(x) = K^{*} (O+F)^{-1} K + B^{-1} \quad (13)$$

We have used (6) to write K instead of Kn'(x). If J(x) is well behaved, then a necessary condition for J to be a minimum is:

$$J'(x)=0 \quad (14)$$

An explicit solution to (14) can be obtained by substitution of (6) and (12). It is instructive for our non-linear problem to do an equivalent derivation using the Newton method. If xi is an approximate minimum of (11) then a better approximation x is given by:

$$x = xi - J''(xi)^{-1} J'(xi) \quad (15)$$

Iteration of this equation should give the desired "maximum likelihood" analysis xa. Since Pof and Pb are Gaussian, and Kn is linear, then J is quadratic, and iteration is unnecessary. Taking xb as the approximate minimum, (15) immediately gives the xa which minimizes J. Substituting (12) and (13) gives:

$$xa = xb + (K^{*}(O+F)^{-1}K + B^{-1})^{-1} K^{*}(O+F)^{-1}(yo - Kn(xb)) \qquad (16)$$

As long as $(KBK^{*} + O+F)$ is non-singular, (16) is equivalent to

$$xa = xb + BK^{*}(KBK^{*} + O+F)^{-1}(yo - Kn(xb)) \qquad (17)$$

This is a matrix form of the general multivariate OI equation; the matrix of weights W being given by

$$W = BK^{*}(KBK^{*} + O+F)^{-1} \qquad (18)$$

In OI a continuous expression is used for the background error covariances, which when evaluated at all grid-points gives B. Kn is regarded as a simple interpolation whose errors are not explicitly considered, so multiplication of B by K is replaced by evaluation of the covariance function at observation positions, and O+F is replaced by a single observational error covariance.

2c. Constrained minimization.

When prior knowledge about possible states x is expressed in probablistic terms, using Pb, then the resulting minimization leads to a weak constraint (Sasaki 1970). If an exact relationship is known then the appropriate probability distribution function becomes a delta function, and it is more convenient to treat it as an additional

strong constraint, either through the use of Legendre multipliers, or by reduction of the control variable (LeDimet and Talagrand 1986). The latter technique can be used if the constraint can be defined by a function Gn relating x to a lower dimensional vector w:

$$x = Gn(w) \tag{19}$$

Then the constrained minimization for x becomes an unconstrained minimization for w. We shall use the prediction model described in section 3 in this way. The space- and time-dimensioned field x is constrained to be exactly consistent with the prediction equations used in the model, so we can reduce the control variable to be the space-dimensioned initial conditions for the prediction model. The penalty function to be minimized with respect to w is:

$$J(w) = -\ln(Pof(Kn(Gn(w))-yo)) \; -\ln(Pb(w-wb)) \tag{20}$$

The background field xb should also obey the constraint, so we have replaced it by wb. Pb now define our prior knowledge about the space-dimensioned fields w.

Although it apparently gives only a space dimensioned result, this technique does find the best four-dimensional field x if the relationship (19) is exact, so that it is appropriate to enforce it as a strong constraint. If (19) is not exact the equations are still valid as long as F takes into account its errors when representing the error with which y can be calculated from w. However the results have to be interpreted differently, since the solution for the best w no

longer gives us immediate knowledge of the best x.

2d. Iterative solutions to non-linear equation.

We now look for iterative methods of finding the minimum of (20) when Kn or Gn are non-linear or Pof or Pb are non-Gaussian. We use wi to denote the current best estimate of wa, the vector which minimizes (20). To simplify notation we use dy to denote the deviations from the observed values of the observation estimates calculated using wi:

$$dy = Kn(Gn(wi)) - yo \tag{21}$$

We also introduce symbols for the components of the total penalty function:

$$Jof(dy) = -\ln(Pof(dy)) \tag{22}$$

$$Jb(x-xb) = -\ln(Pb(x-xb)) \tag{23}$$

Assuming that Gn is differentiable, we can define matrix G by:

$$Gn(wi+dw) = Gn(wi) + G\,dw \tag{24}$$

The penalty function equivalent to (5) then becomes:

$$J(w) = Jof(dy) + Jb(w-wb) \tag{25}$$

Its derivative with respect to w is:

$$J'(w) = G^{*} K^{*} Jof'(dy) + Jb'(w-wb) \qquad (26)$$

If we ignore the dependence of G and K on w, then we can get an approximation for the second derivative:

$$J''(w) = G^{*} K^{*} Jof''(dy) K G + Jb''(w-wb) \qquad (27)$$

If wi is an estimate of the w which mimimizes J, then the Newton method give us a better estimate:

$$w = wi - J''(wi)^{-1} J'(wi) \qquad (28)$$

For Gaussian probability distributions, and using the approximation (27), we get:

$$w = wi - (G^{*} K^{*} (O+F)^{-1} K G + B^{-1})^{-1} (G^{*} K^{*} (O+F)^{-1} dy+B^{-1} (wi-wb)) \qquad (29)$$

Note that for Kn or Gn non-linear, K and G will be functions of wi, whose derivatives we have ignored in (27). This is sometimes called the Gauss-Newton method. Iteration of this equation is the non-linear equivalent of OI, extended to include the time-dimension.

K^{*} and G^{*} are the adjoints of our generalized interpolation Kn and forecast model Gn; their evaluation is in principle straightforward, and requires similar computer resources to the interpolation Kn and forecast Gn. The main restriction is that the adjoint of a non-linear prediction model requires a record of the actual prediction Gn(wi).

Details are given in section 3. Courtier and Talagrand (1986), and Lewis and Derber (1985), give examples of their evaluation in variational analysis methods designed to minimize Jof. The incorporation of background information via Jb makes it necessary also to manipulate the inverse of the covariance matrix B. In Lorenc (1986) it is suggested that this might be done by expressing it in terms of the normal modes of the forecast model (Phillips 1986). Thus the forcing term coming from J'(wi) in (29) can be evaluated. The normalization term, equal to the inverse of J''(wi), requires much more computation, requiring the time-integration of square matrices with the same order as the vector w, and a matrix inversion. This computation requires as many resources as the Kalman-Bucy filter (Ghil et al 1981). Indeed if the forecast model is linear and reversible, (29) is equivalent to the Kalman-Bucy filter. Thus it is worthwhile to investigate methods which do not require the calculation and inversion of J''(wi). Suitable algorithms can be found in textbooks such as Gill et al (1982). The method of steepest descent replaces the normalization term by a scalar, calculated from evaluations of J. The conjugate-gradient method improves on this by using also previous evaluations of J'. Both of these methods converge slowly for eigenmodes of J'' with small eigenvalues. These are likely to be similar to modes of B with large eigenvalues, ie those modes describing the most likely errors in the background field. A possible way of speeding convergence for these important modes is to move a factor B from the normalization term to the forcing term in (29):

$$w = wi - (B\,G\,K^{*}\,(D+F)^{*}\,K^{-1}\,G + I)^{-1}\,(B\,G\,K^{*}\,(D+F)^{*}\,dy+(wi-wb)) \qquad (30)$$

This is equivalent to changing the control variable to v:

$$v = B^{-1} (w-wb) \tag{31}$$

We now have to minimize

$$L(v) = Jof(dy) + Jb(Bv) \tag{32}$$

using

$$L'(v) = B\ G^{*}\ K^{*}\ Jof'(dy) + B\ Jb'(Bv) \tag{33}$$

The Newton method iterative solution for v is then given by:

$$v = vi - (B\ G^{*}\ K^{*}\ (O+F)^{-1}\ K\ G\ B + B)^{-1}\ (B\ G^{*}\ K^{*}\ (O+F)^{-1}\ dy + B\ vi\ ) \tag{34}$$

Descent algorithms applied to (32) and (33) should give faster convergence for the modes where the background is inaccurate, but slower convergence for modes where the background information (such as constraints on unlikely "unmeteorological" modes) is important. It is shown in Lorenc (1986) that if the normalization factor is replaced by a diagonal matrix of sums of unnormalized weights, (30) is a generalization of the successive-correction method.

It is possible in principle to search for the minimum of (25) using a method which does not require the explicit calculation of its derivative (26). So there is no theoretical requirement that Kn and Gn should be differentiable. However in practice the dimensionality

of w is such that any such method would converge extremely slowly.

3. EXPERIMENTAL SYSTEM

In order to study the behavior of the non-linear analysis equations in an ideal situation, we choose a simple system where the prediction model, its adjoint, and the background and observational error distributions can be known exactly. A one-dimensional finite-difference model of the shallow-water equations is used as our strong constraint on the permitted space- and time-values of x. Initialized from smooth initial conditions wt which are known to generate a hydraulic jump, it is integrated forward to generate the truth xt. Background and observational error covariances are defined, and a background wb and observations yo are calculated by adding pseudo-random numbers consistent with these to wt and Kn(xt). Thus all of the results shown in section 4 fall into the class of "identical-twin" experiments.

3a. Forecast model and initial conditions.

The one-dimensional shallow-water equations with rotation are used in the flux form given by Parrett and Cullen (1984 equation 8), who showed that these equations are suitable for simulating a hydraulic jump. The initial conditions wt, taken from their equation 9, are

smooth, but for the parameters chosen develop a hydraulic jump. The equations are integrated, and results are presented, in non-dimensional form. Except where noted the Coriolis parameter and magnitude of the initial perturbation are set to give Rossby and Froude numbers of 1, as in Parrett and Cullen (1984 figure 2). Cyclic boundary conditions are used in space, with resolution 128, and the model is integrated until (non-dimensional) time T=2.8, which is long enough for the hydraulic jump to develop and propagate nearly across the domain.

3b. Background and observational errors.

Phillips (1986) has suggested that a reasonable approximation to the error covariance of a forecast background, in well observed regions, may be obtained by postulating uncorrelated random errors in each of the normal-modes of the linearized forecast model. We use this model of errors to generate the covariance matrix B. By making the energy in each mode a function of frequency, it would be possible in this way to incorporate balance constraints, since with a high-frequency cut-off only the geostrophically balanced modes of the background are assumed to be in error, so only these are changed in the analysis process which uses B. However in our simple example we are interested in an unbalanced gravity mode, so this is inappropriate. Instead we use a simple wavenumber dependency, equivalent to an analysis constaint on smoothness. The error in a mode with wavenumber k is

assumed to be from a Gaussian distribution with zero mean and variance Bk(k) given by:

$$Bk(k) = B0\ (1 + (k/3)^4)^{-1} \qquad (35)$$

This pure wavenumber dependence gives no cross-correlations between the error for different variables in the matrix B. (A high-frequency cut-off would give "geostrophic" cross-correlations between height and transverse velocity). All variables have the same horizontal auto-correlation, shown in Fig.1.

The forward interpolation Kn is performed by calculating the value of the observed parameter at the four surrounding points on the space-time grid used to represent x, and then interpolating linearly between these to the actual position and time of the observation. For these experiments the errors of representativeness F in this forward interpolation are assumed to be zero, and observational errors are assumed to be random and Gaussian, with variance .0025 of that of the same variable in the background.

3c. Adjoint model.

The forecast model Gn is implemented as a sequence of operations of a time-stepping operator Mn upon the latest model states (two of them for the leap-frog time-stepping scheme). Thus its linearization G can be expressed as the product of the linearizations M of Mn about these states, and the adjoint model G^* as the product, in the reverse order, of the time-stepping adjoints M^*. Since the forecast model is

non-linear, its derivatives and adjoints are functions of the model forecast states. The model Gn is integrated from wi to calculate xi. Using these stored values, linearizing the explicit finite-difference time-stepping operator is easy. Two terms are generated for each non-linear product term in the basic model, so the linearized model and its adjoint are between one and two times as complex as the original model. As explained by Talagrand and Courtier (1986), calculation of the term $G^* K^* Jof'(dy)$ in (26) can be done by a single integration, backwards in time, of the adjoints M^* of each time-step of the forward model, accumulating the forcing terms $K^* Jof'(dy)$ at their valid times. The normalization term $G^* K^* Jof''(dy)KG$ in (27) can be similarly treated, however for this a matrix rather than a model state needs to be integrated backwards in time, approximately squaring the number of computations necessary each time-step.

4. RESULTS AND METEOROLOGICAL INTERPRETATION.

In this section we present and discuss optimal non-linear analyses obtained by minimizing (25) for the system described in section 3. Our object is to demonstrate what is theoretically achievable in an ideal situation. Discussion of more practical matters such as the convergence rate of various methods, and approximations, is postponed until section 5. Results are discussed in 5 sub-sections, corresponding to the 5 meteorological analysis problems listed in the introduction. The experiments performed are listed in Table.1 and Table.2.

4a. Four-dimensional data assimilation

Four-dimensional data assimilation can be defined to be the use of a four-dimensional distribution of observations, together with a constraint on the resulting four-dimensional analysis that its evolution in time should satisfy known prognostic equations, as embodied in a forecast model (Lorenc 1986). Since forecast models are not perfect, the latter constraint should not be strictly enforced. Practical implementation of such a scheme is severely limited by available computer resources. The storage and manipulation of high-resolution four-dimensional analyses requires many more resources than running an NWP forecast model, which only manipulates three-dimensional fields. Research in this area has had to severely limit the complexity of the forecast constraints, and the vertical and time resolution of the analysis (e.g. Lewis and Bloom 1978).

The traditional NWP method of approximating four-dimensional data assimilation is the analysis-forecast cycle, in which observations are inserted using a purely three-dimensional analysis procedure into a "background" forecast from the results of the previous analysis. This method does allow for the imperfections of the forecast model; an estimated background error is used to calculate the observation weights in statistical analysis methods such as OI. However it does not properly use time-tendency information in the observations, since observations from different times are not analysed together.

If the prognostic equations are applied as a strong constraint as in section 2b, then some of the computational difficulties can be avoided

through a reduction of the control variable. The four-dimensional analysis is defined by a three-dimensional initial field plus the prognostic equations, and (except for the current estimate xi) four-dimensional fields need not be stored and manipulated, while still obtaining an optimal four-dimensional use of observational information. However this does not allow for inaccuracies in the forecast model over the period of the observations used.

This work combines the latter "strong constraint" approach for a short time-period, with the traditional analysis-forecast cycle approach, so that time-tendency observations within this short period can be used optimally, while information from earlier observations can be used without assuming that the forecast model is perfect. Thus while it is not truly optimal in its application of imperfect prognostic constraints, it does match well the operational NWP requirement of providing initial conditions for a forecast, using the last analysis and observations valid during the period since the last analysis was made.

Fig.2 shows the height field at time T=0 from experiments to illustrate this. In this and most subsequent figures the fields are plotted as solid lines, the "truth" from which observations were generated as dashed lines, and observed values as asterisks whose height covers six times the observational error standard deviation, and whose width covers one model gridlength. The bottom curve in Fig.2 shows the background field, which in practice would come from a forecast from earlier analyses. The bottom curve in Fig.3 and Fig.4 shows a forecast from this background at T=1.4 and T=2.8, with the

corresponding "true" field for comparison. To simulate the current operational situation, with most observations at main synoptic hours, we assume a uniform space-distribution of observations at T=0 and T=1.4. For simplicity of presentation we only use height observations for the first experiments. The curves labelled A in Fig.2 and Fig.3 show the optimal non-linear analysis obtained using these observations. Fig.4 shows forecasts valid at T=2.8 which can be used to judge the extent that the information has been assimilated into the model. Experiments B and C together used the same observations in an analysis-forecast cycle. The curve labelled B in Fig.2 shows the optimal analysis using only the observations available at T=0. With the covariances we have assumed, this is equivalent to a univariate OI of the height field at T=0, not altering the background winds. Forecasts from this are clearly inferior to experiment A. Experiment C used the forecast valid at T=1.4 as background for an analysis, in space only, of the observations available at this time. This is also clearly inferior to experiment A. At T=2.8 the hydraulic jump is as badly positioned as in the background, while experiment A has both the position and shape more nearly correct.

For some modern observing systems, observations are not at all synoptic, but rather spread evenly in time. Satellite temperature soundings were the first important example of this: an important example in the future will be fixed "profilers", giving a detailed time-history of the atmospheric profile at a few horizontal locations. Experiment T demonstrates that our method can make optimal use of such observations. To simulate a single profiler we generate u- and

v-momentum observations equally distributed in time at a point in the middle of our horizontal domain. Fig.5 shows the u-momentum observations, and the time-evolution of the analysis at that point made using them. We can see clearly, near T=1.4, the passage through the observation position of the hydraulic jump. Since it is a propagating system it should be possible to deduce the horizontal structure of the jump near the profiler from this information. Examination of curve T in Fig.3 shows that this has indeed been done.

4b. Use of "indirect" observations.

Early experiments in four-dimensional data assimilation used rather ad hoc modifications of the "direct insertion" method, whereby observations were inserted into the forecast model state at and near their valid position and time. Using this method it is not clear how to use observations not directly related to model variables; observations are transformed to model variables before being inserted. Thus for instance observations of outgoing radiances from satellite temperature sounders are usually inverted into temperatures before being assimilated. In section 2 the objective analysis is itself couched as an inverse problem, with the transformation Kn from model space to observation space explicitly considered, so prior inversion is no longer necessary. As long as estimates of the observed parameters can be calculated from the model using a well-behaved differentiable function Kn, the observations can be used in the analysis. Thus for radiances Kn would be the radiative transfer equation, calculating the outgoing radiation from the model's

temperature and humidity profile. This integrates a non-linear physical-statisical retrieval algorithm for the satellite temperature profiles (Rodgers 1976) into the four-dimensional data assimilation. As NWP models are incorporating more sophisticated physical parameterization and output schemes, the potential for using indirect observations in this way is increasing. For instance observations of outgoing longwave radiation (OLR) from convective clouds (Julian 1984) could in principle be used, with Kn incorporating the model's parameterization schemes for convection and radiation.

Another satellite instrument is capable of deducing the surface wind speed over the ocean from microwave observations, but not its direction. Experiment S simulates such an observing system. We assume that observations are of parameter s, which can be calculated from the basic model parameters height h, and momentum components U and V by:

$$s = (U^2 + V^2) / h \qquad (36)$$

This, together with the space- and time-interpolation from the nearest model points, constitutes our generalized interpolation operator Kn. It is clearly non-linear, particularly near the hydraulic jump where h can become very small. However its differential K is well defined, since h is constrained to be always greater than zero, and an optimal analysis should exist. Observations of s are distributed for experiment S at the same positions as the h observations of experiment A. Fig.6 and Fig.7 show that the analysis from experiment S fits these data closely, more so than that from

experiment A. Although the experiment S analysis does not reproduce the detailed structure of the jump at T=1.4 as well as experiment A, it does give a better prediction of the jump's position and structure at T=2.8 (Fig.4).

Another type of data, used in subjective analyses and forecasts, but not useful in conventional updating data assimilation methods, is tracer information. A time-series of observations of a parameter advected by the wind field provides information about the wind field as well as about the advected field. We can simulate this in our simple model by setting the Coriolis parameter to zero, uncoupling V from h and U, and using V as a tracer. Fig.8 and Fig.9 show experiments with this system. In experiment VA, the full adjoint technique is used to modify h, U and V to obtain the best fit to the V observations. In experiments VB and VC, h and U are kept at their background values, and an analysis-forecast cycle (as in experiments B and C) is used to update the V field. Useful corrections to the advecting velocity are made in experiment VA, so that its forecast V field at T=2.8 is better than that of experiment VC (Fig.10).

4c. Analysis of discontinuities and fronts.

For many years human analysts have had conceptual models of fronts, and have fitted these to rather sparce data, producing analyses with detailed structure such as sharp windshears, even when these were not well resolved by the observations. This process is non-linear; the

use the analyst makes of an observation depends on what he believes the meteorological situation to be, based in part on the observation value. Attempts to formalize such conceptual models and automate this analysis process have not been very successful. However we now have high-resolution NWP models which generate naturally very realistic looking frontal structures. These forecast models might be able to replace the conceptual models in a non-linear analysis scheme, to produce analyses which are consistent both with the observations and with the model's dynamics. Given a background with a front somewhat misplaced, a human analyst will move it to fit available observations, while keeping basically the same structure. A linear analysis scheme like (1) cannot do this, as illustrated in Fig.3 curve E, which shows a linear, space-only, analysis of the observations at T=1.4 using the background shown in the bottom curve. The linear analysis has put a jump near the correct position, however its structure is incorrect, and "shadows" remain of jumps in incorrect positions in the background field. This can be compared with the non-linear analysis shown in curve D, which used the same observations, but did a space- and time-analysis using as background the bottom curve in Fig.2. The non-linear method has moved the hydraulic jump at T=1.4 to fit the data, while keeping its structure. Indeed this analysis is better than that from the analysis-forecast cycle (C), which used, linearly, twice as many observations. The beneficial non-linear used of observations in experiment D comes from the used of the non-linear forecast model to link T=0 with T=1.4. We still assume that errors at T=0 in wb are described by B, that is that they are large scale and there is no non-linear coupling of scales. The method only gives an improvement in the analysis at T=1.4 insofar as the small scales at

that time are determined by the larger scales at T=0. This is true for our chosen example; it is also probably true to a large extent for atmospheric fronts, whose positions and structure depend on the larger scale forcing.

4d. Balance constraints, initialization.

We discussed in section 3b how, by specifying the background error for each mode as a function of its frequency, we could incorporate a geostrophic balance constraint on the deviations of the analysis from the background. Such a constraint is linear, and only constrains the total analysis field if the background is itself balanced. It is sometimes desirable to incorporate a non-linear constraint on the total analysis into the analysis process. Such a constraint can be justified if we have prior knowledge that the atmosphere is usually slowly varying. NWP models, if integrated for a sufficiently long time, also have this property. However since we are only using the model as a strong constraint over a short time-period, for which the model will also allow rapidly varying gravity-wave modes, we must use other means to enforce the constraint. The simplest way is to add to the total penalty (25), a penalty term Jin(w) which penalizes rapid variations. In order to have the same effect as non-linear initialization, we implement this penalty on the time derivatives calculated by the NWP model during its first timestep at T=0.

$$Jin(w) = (w-Mn(w))^{*} C (w-Mn(w)) .5/dT^{2} \qquad (37)$$

C is a matrix specifying the penalty on each model variable. Since we are already calculating $M_n(w)$ and M^* during the analysis, the addition of this penalty and its derivatives to (25) (26) and (27) is easy. Experiment TIN incorporated such a penalty, with other details identical to experiment T. In current operational schemes a non-linear initialization is often applied after the analysis is complete. This can significantly decrease the closeness of fit to the observations. We see in Fig.5 that experiment TIN has fitted the observations as closely as experiment T, while reducing slightly the initial rapid variation in U. Errors relative to the "true" field still exist in the large scale h field, since this is defined neither by the observations nor by the constraint (Fig.3), but the spurious jumps introduced by the background, and not altered except in position in experiment T, have been reduced by the initialization constraint. This causes the forecast structure at T=2.8 to be better (Fig.4).

4e. Non-Gaussian observational errors and quality control.

Unfortunately the observations available for routine NWP occasionally deviate by a large amount from the true value, because of gross error, either human, or in the instrument or communication system. Many more such errors occur than would be expected from the Gaussian distribution which describes the majority of errors. Traditionally such data are searched for and eliminated during a preliminary quality control step. Purser (1984) suggested that an alternative approach would be to consider the non-Gaussian error distribution directly in

the analysis. A simple model for the distribution of gross errors has been put forward and tested by Lorenc and Hammon (1986). They postulated that there was a small probability of a gross error event occuring, and if it did occur the observed value had no useful information, but was equally likely to be any value within a range of plausible values as defined by (say) several times the background error standard deviation. This leads to a probability distribution function for a single observation [i], from both instrumental and gross errors, valid in the range of plausible values:

$$Po(y[i]-yo[i]) = \exp(-(y[i]-yo[i])^{*} \; O[i]^{-1} \; (y[i]-yo[i])/2) + g \qquad (38)$$

Here yo[i] is a single observed value, O[i] is its instrumental error variance, and g is a constant depending on the prior probability of a gross error and the range over which gross errors are distributed. Again we have omitted normalization constants, as in (8). If the occurences of gross errors in different observations are independent, like the instrumental errors, then the combined probability distribution for many observations will be the product of these factors. Taking logarithms then gives (ignoring the convolution with errors of representativeness):

$$Jof(dy) = \sum_{i} -\ln(\exp(-(y[i]-yo[i])^{*} \; O[i]^{-1} \; (y[i]-yo[i])/2) + g) \qquad (39)$$

If the probability of a gross error occuring is .05, and observations in gross error give values equally spread within four standard deviations of the background value, then with the background and

observation errors we are using, g equals 0.000825. Fig.11 shows the penalty function for a single observation using these values (shifted by ln(g)), with the equivalent quadratic penalty function for a pure Gaussian distribution for comparison. Near the observed value the functions are identical. However farther away the new function asymtotes to a plateau value, rather than continuing to increase. This has very important consequences for the total penalty function, since it makes the existence of multiple local minima much more likely. Our iterative mimimization algorithms are designed to find any minimum, rather than the smallest, so they are therefore much less likely to find the most likely analysis.

Experiments illustrating this are shown in Fig.12 and Fig.13, and listed in Table.2. Particularly when using the non-Gaussian observational error probability distribution (38), the minimum located by the iterative solution method depends greatly on the first guess used to start the iteration. Thus experiment AQ fitted very few of the data, finding a local minimum near to its first guess, which was the background. This can be compared with experiment AQA, which found a lower mimimum near the analysis of experiment A, fitting all the data. When one of the data values was arbitarily increased by 0.5, to simulate a gross error, then starting from experiment A as first guess the datum was rejected while others were fitted (experiment AQGA in Fig.13). This is the result we might hope for from an ideal scheme. However the analysis (not shown) starting from the background field, fitted as few data as experiment AQ. Moreover the "bad" datum we generated was not completely implausible; experiment AG fitted it by moving the nearby peak in the background field. Starting with this

analysis as first guess, experiment AQGAG continued to fit the "bad" datum, rejecting its neighbours instead. The minimum found in experiment AQGA was lowest of those shown, but a more complex search algorithm would be necessary to be sure it was the absolute minimum. Problems of multiple minima are discussed further in section 5c. It is clear however that simply using a non-Gaussian penalty function does not relieve us from the need for complex logic and decision taking algorithms associated with traditional quality control methods.

5. SOLUTION METHODS.

In section 4 we were concerned with the "meteorological" properties of the solutions obtained, without regard to the means of obtaining them. In this section we discuss practical problems of solving the equations. So as to be able to continue iteration until we are sure a minimum has been found, and to perform the matrix manipulations required to calculate and invert the normalization factor J''(wi), experiments were performed with a horizontal grid of 64 points. They are listed in Table.3. After describing, in section 5a, an ad hoc additional constraint which was found to be necessary, we go on to describe the descent algorithms tried and their properties in section 5b. In section 5c we discuss some examples of multiple minima, and in section 5d some possible practical approximations.

5a. Constraint on realistic heights in forecast model.

The forecast model Gn, which we use to calculate the space-time fields xi from the initial space field wi, has truncation errors in its time-stepping algorithm and can generate negative values for the height h. This unphysical result must be avoided, however the "correct" way of doing so, reducing the time-step, is not easy to implement in practice. Instead we choose time-step and diffusion coefficients such that neither the "truth" nor the background give h less than 0.1, and keep these values for all experiments. When, as quite often occurs, a new estimate of wi gives heights less than or equal to zero, then the integration of the forecast model is stopped, and the iteration is restarted using a modified wi formed by averaging wi with that giving the last successful integration. In order to avoid this abrupt limit being a factor in any final solution, an additional penalty function Jh(wi) is included in the total penalty J(wi) given by (25):

$$Jh(wi) = Ch \sum_{grid} \max(0, hl-h)^3 \qquad (40)$$

Here hl is the arbitrarily chosen limit 0.1, Ch is a constant coefficient, and the summation is taken over all h values in Gn(wi). If this penalty is large the iteration is restarted, as described above. Otherwise it and its derivatives are added to (25) (26) and (27). This has little effect on the final solutions for the examples shown; indeed for the "truth" and the background it has no effect, since these have no h values less than hl. However restarting the

descent algorithms with modified wi, as described above, has considerable effect on their speed of convergence. The additional penalty (40) also slows convergence to fitting the observations of small values of h, like those in the centre of Fig.2, although in most cases a close fit is finally achieved, as can be seen in the figure.

5b. Descent algorithms, and speed of convergence.

In OI the penalty function J is quadratic, and the optimal analysis is found in one iteration by using (17). The matrices which have to be inverted are kept small by truncating the data selected to influence each grid-point. Courtier and Talagrand (1986) performed experiments with a simple vorticity equation model and sufficient observations to determine the solution without using any background information (although they did find useful a constaint that the analysis be smooth). They demonstrated that their penalty function was approximately quadratic and iterated using (25) and (26) (excluding the background terms), with spectral coefficients as control variable. They did not thoroughly investigate descent algorithms, however they stated that the steepest descent algorithm did not converge in a reasonable number of iterations, the conjugate-gradient algorithm gave acceptable results after between 10 and 100 iterations, and a quasi-Newton method was more efficient. Bratseth (1986) has presented an iterative successive-correction method which is equivalent to OI; the method is similar to (30) (see Lorenc 1986 for more details). In experiments with this scheme Gronas and Midtbo (1986) found that when there are no serious conflicts

between the error statistics and the observations, 5-10 iterations may be sufficient. Slower convergence resulted for modes where the background was less likely to be in error, and was often associated with erroneous observations.

In this work the penalty function J can be far from quadratic for any of the reasons listed in the introduction. It is therefore of interest to see whether a reasonable speed of convergence can be achieved, as guidance as to whether it might be feasible to include such non-linear effects in an operational analysis system. The rates of convergence of five different descent algorithms are compared in Fig.14. Experiment 1A, shown as a solid line, used the Newton method of (29). For a penalty function which is approximately quadratic this should give a rapid convergence, the fact that it did not demonstrates the importance of the non-linearity of Gn. For the first few iterations convergence was also slowed by the additional constraint on h being positive, as discussed in section 5a. Integration of the adjoint model for the matrix term in Jof''(wi) is extremely expensive. In experiment 1AM (shown as long dashes) an attempt was made to save time on this by only recalculating it every tenth iteration. This converged significantly slower, another indication the J is non-quadratic. However, despite this, the modified method required less computation for a given convergence. Experiment 2A, shown as a dotted line, used a conjugate-gradient algorithm (Powell 1977). If not restarted by the procedure of section 5a, the method was restarted every twentieth iteration, the first step after each restart being in the direction of steepest descent. This method only uses the penalty

(25) and its gradient (26); it is much less costly in both computer time and storage requirements than the first. Experiment 3A, shown as short dashes, was like experiment 2A, but using as control variable the transformed variable v (31). Finally in experiment 4A, shown as a dash-dot line, methods 3 and 2 were alternated each restart. This method was concidered to be the most robust, and was used for the experiments shown in section 4. The final states achieved by each method were nearly identical; that from method 3 is shown in Fig.18 Fig.19 and Fig.20 curve 3A. Differences were probably due to computer truncation errors. Method 1 is in fact searching for the zero of the gradient J'(w), while the other methods are searching for the minimum of J(w). Method 3 requires extra multiplications by B. It was found that double precision computations were desirable for close convergence of methods 2 3 and 4, while they were essential for calculation of the matrix Jof''(wi) in method 1.

For this example, the most cost-effective method of achieving final convergence is method 2. However this in practice is not the only criterion for choosing a method. The final stages of convergence can be making detailed changes well within the uncertainty about what is the "true" state. This is shown in Fig.15 which shows the convergence rates for experiments like those shown in Fig.14, but starting from a first guess equal to the "true" state from which the background and observations were generated. Method 3 can be seen to be converging extremely slowly, but this is of no practical importance. If we look at Fig.14(b) we see that on the way to the final converged state method 2 gives large values of the background penalty Jb; nearly three time that of the "true" state (which of course in these idealized

experiments we know). This ratio should be between 0 (if the observations are useless), and 1 (if the observations alone are sufficient to define the "true" state). All except method 3 seem to achieve convergence by first fitting the observations closely, then modifying the state reached to comply with the background constraint, which in our case was simply on the smoothness of the deviation from wb. This effect is illustrated in Fig.16 for a simpler, space-only example in which fewer h observations, at the single time T=0.0, were analysed. In this case wi is equivalent to xi, and we are finding the minimum of the quadratic penalty (11). Method 1 is a form of OI, and converges in one iteration. Method 3 actually converges slightly faster than method 2 in this example, but there seems little to choose between their convergence rates for the total penalty. However we see in Fig.16(b) that there is a big difference between their convergence for the background penalty. Method 2, until it has nearly fully converged, has large values of the background penalty. This is equivalent to saying that it gives states which our prior knowledge leads us to believe are unlikely to be correct. If we have confidence in our prior knowledge this is clearly undesirable. In this example our prior knowledge only constrains the smoothness of the h fields' difference from the background. Fig.17 curves 2B and 3B show the states given by methods 2 and 3 after five iterations. Curve 2B is undesirably rough. Thus if we are looking for a method to be used practically with a limited number of iterations, and if our prior knowledge is more important than a close fit to all observations, method 3 is to be prefered.

5c. Multiple minima.

The convergence properties of any method become greatly different from those for a quadratic penalty function when the penalty has multiple minima, or regions with small or zero gradient. We saw in Fig.11 how an observation which has a finite chance of being incorrect gives a penalty function which has nearly horizontal plateaus away from the observed value. In these regions the local derivative of the penalty function does not give any information about which direction to search for a lower value. Minor perturbations to the penalty, as can easily occur in a problem including a model with wave-like solutions, can convert such plateaus to local minima. In such cases solution methods can converge to a local minimum. Which minimum is located depends on the initial guess for the iterative algorithm, as illustrated by the experiments discussed in section 4e. It also depends on which descent algorithm is used; experiments 2AQ and 3AQ use methods 2 and 3 repectively. Their analyses (Fig.17 and Fig.18) show no sign of further convergence after 800 iterations each. Method 3 has found the lower minimum, with penalty equal to 174.5 compared to method 2's 313.0. Multiple minima are not confined to the non-Gaussian observation error distribution experiments in our examples. If the number of data is reduced slightly, so that the shape of the jump at T=1.4 is even less well resolved, then a solution exists with the wrong peak of the wave fitting the maximum observation. Experiment 3AO in Fig.18 illustrates this. A penalty function with plateaus would also result from the use of indirect data, such as OLR from convective clouds, which relate to processes which in the forecast model have distinct limits for their initiation.

If the current best estimate wi does not imply convection, then even if the forward process Kn(Gn(wi)) is differentiable, it is unlikely that its gradient will indicate in which direction to move to initiate convection. As with all non-linear systems, these problems can be reduced by having as good a first estimate as possible, so that linearization about it is accurate over the range of likely states.

5d. Approximations.

Practical analysis schemes for NWP need to process very large numbers of observations and high resolution forecast models, and the results must be available within strict time-constraints. One purpose of this work is to study approximations to the ideal equations, and their cost-effectiveness. The first approximation is unavoidable. Actual forecast models are not perfect, and the assumption that the forecast model can be used as a strong constraint needs to be tested. We do this by performing experiment 3AT in which the "truth" is generated by a model which is not an identical-twin of the model used for the assimilation. The observations from experiment A are analysed using a model with half the resolution. The results are shown in Fig.17 Fig.18 and Fig.19 curve 3AT. They can be compared with curve 3A, which shows the equivalent identical-twin experiment, and with curve A, which was interpolated from the higher resolution identical-twin experiment A. Experiment 3AT did successfully position the hydraulic jump at T=1.4 (Fig.18), however it could not get the correct wavelength for the ripples behind the jump; these are a function of

the model's resolution. The forecast for T=2.8 (Fig.19) was not as good as in either of the identical-twin experiments, but still had some value.

The effect of the non-linearity of Gn, making (24) an inaccurate approximation, can be demonstrated by reducing all error variances by a constant factor. If this is done consistently both to the variances in B and O+F, and to the actual pseudo-random errors used when generating observations and background from the "truth", then (to a first approximation) the initial and final value of the penalty function are not altered. However dw in (24) is altered, and (24) becomes a better approximation. Experiment 3AS is like experiment 3A but with error variances reduced by a factor of 10. The resulting analyses are shown in Fig.17 Fig.18 and Fig.19, and the rate of convergence in Fig.20 (solid line). The experiment does as expected converge more rapidly than experiment 3A (dashed line) especially initially when the non-linear effect discussed in section 5a slows convergence of experiment 3A.

The computation of J''(wi) using (27) is impracticable for a model with realistic resolution. Moreover a manipulation analagous to that giving (17), which reduces the matrix order to the number of data used, is not available to us if we wish to use the adjoint model technique to reduce the control variable in the time-dimension. We are forced therefore to use a minimization algorithm which approximates or does not need this. Lorenc (1986) explains how the successive-correction method can be thought of as using (30) with the term coming from J''(wi) approximated by a diagonal matrix of the sum

of unnormalized weights. In this work we have used the conjugate-gradient method. Research into efficient approximate descent algorithms combining these approaches is in progress.

It is important for practical schemes that iteration can be truncated before complete convergence. We saw in section 5b that method 3, using (31) (32) and (33), always gives a state which is not too unlikely given our prior background knowledge. It might not, however, fit all the observations closely. These are desirable properties for a NWP analysis scheme, since observations occasionally have gross errors. Another advantage of approximations based on method 3 is that the inverse of B is not needed, and B can be singular. For instance if we wish to assume that the errors in wb are precisely geostrophic, we can set to zero the error variances in ageostrophic modes, identified in section 3b as those modes with non-zero frequency. Such error covariance models are routinely used in OI. We see in Fig.14(b) that method 3 converges extremely slowly for those modes which contribute most to the background penalty. If, as is usually done, we start the iteration with wi=wb, then initially the background penalty will be zero. This suggests that a further approximation can be made by ignoring the background penalty, if we are going to truncate the iteration before complete convergence. Curve 3AG of Fig.17 Fig.18 and Fig.19 show the results of such a scheme after 30 iterations. The major features of the analysis are correct. The dotted curves of Fig.20 show the convergence of this approximate scheme. Iterated indefinitely the scheme will, if it is possible, fit the observations exactly, something which is undesirable for our inaccurate

observations and useful background information. But after a finite number of iterations this will not be the case. Talagrand and Courtier (1986) have made such an approximation, it is also implicit in the old successive-correction methods (Lorenc 1986).

Another possibility for approximation is available as long as the analysis is regarded as the minimization of the penalty (25), rather than finding the zero of its derivative (26). As mentioned at the end of section 2, the derivative is then only needed to speed convergence in the search for this minimum, by indicating a search direction. For this it might be possible to approximate the adjoint model $\mathbf{G}^*$. Approximations of this sort have yet to be sought for this system.

6. DISCUSSION

We have presented in section 2 a derivation of analysis equations which can be non-linear in the data values. That is, the weights given to observations depend on the analysed fields, not just their positions and accuracy. The equations can be used when we have non-linear relationships constraining the analysis, or when the observed data are non-linearly related to the desired fields, or when the error distributions are non-Gaussian. They are optimal, if we

assume that the best estimate is that which is most likely. For the linear Gaussian case they simplify to the well known OI equations.

Simplifications can be made in the use of prognostic constraints if we assume that, for a short period, errors in a forecast model are negligable, so it can be used as a strong constraint in space and time on the allowable solutions. This reduces the analysis to a space only problem, with the integration of the forecast model and its adjoint, as suggested by LeDimet and Talagrand (1986). Iterative solution methods for this system are derived.

In order to test the behavior of the scheme in non-linear cases, we set it up for a shallow-water model which forecasts the evolution of a hydraulic jump, described in section 3. Examples analagous to practical problems in objective analysis for numerical weather prediction were presented in section 4. We demonstrated that:-

a. The non-linear analysis method is able to use time-tendency information from observations better than the analysis-forecast cycle method which is normally used for data assimilation. It could also convert a time-sequence of data from a single observation location into useful information about the spatial structure of the field.

b. The data used in the analysis do not have to be transformed to the analysed parameters; it is only necessary to have a known method for calculating the observed parameters from those analysed. Thus observations of wind speed, without direction information, can be used, as can information about the advection of a tracer.

c. If the forecast model generates realistic structures for features

like fronts, then the non-linear method is capable of "moving" such a feature in the first-guess to fit the available data, even if the data do not resolve all the details of the feature. The resulting analysis is thus more detailed than a scale analysis of the observational distribution alone would lead one to expect.

d. An additional constraint, that the non-linear evolution of the analysed field should be slow, can be incorporated as part of the analysis process. This improves the evolution of the subsequent forecast from the analysis, without greatly reducing the fit of the analysis to the observations.

e. Observations which are more likely to have large errors than would be expected from a normal distribution of instrumental errors, can be allowed for by specifying an appropriate non-Gaussian error distribution. If we make the reasonable assumption that observations with such gross errors contain no useful information, then a limit is placed on the penalty function being minimized. This generates "plateau" regions, and greatly increases the difficulties in finding the best analysis if we do not have an accurate first-guess. Thus the complex logic required for a comprehensive quality control is not avoided. However, given a reasonable first-guess, the method does effectively ignore erroneous data.

The properties of the methods used to solve the non-linear analysis equations were studied in section 5, with a view to providing guidance as to possible methods for practical schemes. The Gauss-Newton method, which is equivalent to OI for the linear case, was found to be very expensive when used for the data assimilation problem. When the

analysis was significantly non-linear its convergence was often rather slow. Approximations to the normalization factor slowed convergence even more. Experience with the successive-correction method, which uses a very inexpensive approximation to the normalization factor, suggests that it may be useful to seek a similar approximation for these equations. Research in this area is continuing. Other experiment studied the conjugate-gradient method. The choice of control variable for this descent algorithm was found to be crucial for the characteristics of the analysed fields if iteration was not taken to convergence. Use of a grid-point representation, and a straightforward application of the conjugate-gradient algorithm to the basic equations (25) and (26), gave fields after a few iterations that were unlikely, given our prior knowledge of the background field. This is undesirable for practical schemes. By transforming the control variable using the background error covariance, and applying the conjugate-gradient method to (32) and (33), we could ensure that the iteration always gave fields which were not unlikely. Because this algorithm ensures that the background penalty remains small, we could make a further approximation by ignoring it.

For cases where the analysis was so non-linear that the penalty function had multiple minima, the importance of a good initial estimate was demonstrated. For observations with a possibility of gross errors a preliminary analysis using data which passed a preliminary quality control could be used. For non-linear indirect data such as observations of OLR, the process observed, convection, should be predicted in the first guess. Otherwise it is unlikely that

the gradient of the penalty will indicate in which direction to search to initiate convection. It should be noted that the initial estimate need not be the background field for (29) to be valid, nor need it be independent of the observations to be used. However if it does not give a low value for the background penalty Jb, then the advantages of using the transformed control variable v and a truncated number of iterations will be lost.

The iterative analysis method studied in this work is not the only way of introducing non-linear effects into the basic, linear, methods such as OI. Some effects can be allowed for by determining some of the "constant" coefficients used in the linear method from a preliminary scan of the observations and background. One example is the use of background error covariances which are dependent on the local meteorological situation, another is the use of a preliminary quality control scan to detect and reject erroneous data.

This work was performed while the author was a visiting scientist at the NOAA, NWS, National Meteorological Center, Washington, DC.

REFERENCES

Bratseth, A.M. 1986 "Statistical interpolation by means of successive corrections." Tellus, 38A, to appear

Courtier, P., and Talagrand, O. 1986 "Variational assimilation of meteorological observations with the adjoint vorticty equations. - Part II. Numerical results." Quart. J. R. Met. Soc., submitted

Ghil,M., Cohn,S.E., Tavantzis,J., Bube,K., and Isaacson,E. 1981 "Applications of estimation theory to numerical weather prediction." Dynamical meteorology: Data assimilation methods. eds. Bengtsson,L., Ghil,M., and Kallen,E. New York, Springer-Verlag, 139-224

Gill,P.E., Murray,W., and Wright,M.H. 1982 "Practical optimization" Academic Press. London.

Gronas,S., and Midtbo,K.H. 1986 "Four dimensional data assimilation at the Norwegian Meteorological Institute" Norwergian Met. Inst., Oslo. Tech. Rept. No 66. 66pp

Hoffmann,R.N. 1986 "A four dimensional analysis exactly satisfying equations of motion." Mon. Wea. Rev., 114, 388-397

Julian,P.R. 1984 "Objective analysis in the Tropics: A proposed scheme." Mon. Wea. Rev., 112, 1752-1767

LeDimet,F-X., and Talagrand,O. 1986 "Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects." Tellus, 38A, 97-110

Lewis,J.M. and Bloom,S.C. 1978 "Incorporation of time continuity into subsynoptic analysis by using dynamical constraints." Tellus, 30, 496-516

Lewis,J.M. and Derber,J.C. 1985 "The use of adjoint equations to solve a variational adjustment problem with advective constraints." Tellus, 37A, 309-322

Lorenc,A.C. 1986 "Analysis methods for numerical weather prediction." Quart. J. Roy. Met. Soc., 112, ???-???

Lorenc,A.C., and Hammon,O. 1986 "Objective quality control of observations using Bayesian methods - Theory, and a practical implementation." Met O 11 Tech. Note 225.

Parrett,C.A., and Cullen,M.J.P. 1984 "Simulation of hydraulic jumps in the presence of rotation and mountains." Quart. J. R. Met. Soc., 110, 147-165

Phillips,N.A. 1986 "The spatial structure of random geostrophic modes and first-guess errors." Tellus, 38A, ???-???

Powell,M.J.D. 1977 "Restart procedures for the conjugate gradient method." Mathematical Programming, 12, 241-254

Purser,R.J. 1984 "A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis." Preprints, 10th conference on weather forecasting and analysis. Am. Met. Soc., 102-105

Rodgers,C.D. 1976 "Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation." Rev. Geophys. Space Phys., 14, 609-624

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

Talagrand,O., and Courtier,P. 1986 "Variational assimilation of meteorological observations with the adjoint vorticity equation - Part I. Theory." Quart. J. Roy. Met. Soc., submitted.

Wahba,G. 1982 "Variational methods in simultaneous optimum interpolation and initialization." The interaction between objective analysis and initialization: Proc, 14th Stanstead seminar. ed Williamson,D., Publ. in Meteorology 127, McGill Univ. Montreal., 178-185

Williamson,D., and Daley,R. 1983 "A unified analysis-initialization technique." Mon. Wea. Rev., 111, 1517-1536

Table 1. Experiments described in sections 4a to 4d.

Curves in Fig.2 to Fig.10 are labelled with Expt.

Observations were either equally distributed in space at each time, or equally distributed in time at one position.

Experiments used a space grid of 128 points.

Expt	Observations	Description
A	35 h at T=0.0 35 h at T=1.4	baseline optimal non-linear analysis.
B&C		analysis-forecast cycle:
B	35 h at T=0.0	linear analysis at T=0
C	35 h at T=1.4	linear analysis at T=1.4, using a background forecast from expt B.
D	35 h at T=1.4	optimal non-linear analysis.
E	35 h at T=1.4	linear analysis at T=1.4, using a background forecast from the T=0 background.
S	35 s at T=0.0 35 s at T=1.4	optimal non-linear analysis of observations of wind-speed squared s = (U*U + V*V)/h*h.
VA	35 V at T=0.0 35 V at T=1.4	optimal non-linear analysis of observations of a tracer V (with Coriolis parameter = 0).
VB&VC		analysis-forecast cycle:
VB	35 V at T=0.0	linear analysis at T=0
VC	35 V at T=1.4	linear analysis at T=1.4 using VB as background, of a tracer V (with Coriolis parameter = 0).
T	38 U at S=pi 38 V at S=pi	optimal non-linear analysis.
TIN	38 U at S=pi 38 V at S=pi	optimal non-linear analysis, with additional penalty on rapid variations.

Table 2. Experiments described in sections 4e.

Curves in Fig.12 & Fig.13 are labelled with Expt.

35 h observations were equally distributed in space at T=0.0 and at T=1.4, as in experiment A.

Experiments used a space grid of 128 points.

Expt.	Pof	erroneous data: First-guess
A	Gaussian	0 : background
AQ	non-Gaussian	0 : background
AQA	non-Gaussian	0 : analysis from A
AG	Gaussian	1 at T=1.4 : background
AQGA	non-Gaussian	1 at T=1.4 : analysis from A
AQGAG	non-Gaussian	1 at T=1.4 : analysis from AG

Table 3. Experiments described in section 5.

Curves in Fig.14 to Fig.17 are labelled with Expt.

Observations were distributed as in A (Table.1) except as noted.

Experiments used a space grid of 64 points.

Expt	Iteration method and modifications to parameters.
1A	Newton.
1AM	Newton, modified to not recalculate inverse of J''(wi) each iteration.
2A	Conjugate-gradient, with restarts.
3A	Conjugate-gradient, with restarts, transformed control variable (v).
4A	Conjugate-gradient, with restarts, alternating control variables (v & w).
1B	As 1A. 10 h observations at T=0, analysed in space only.
2B	As 2A. 10 h observations at T=0, analysed in space only.
3B	As 3A. 10 h observations at T=0, analysed in space only.
2AQ	As 2A. Non-Gaussian observation penalty function.
3AQ	As 3A. Non-Gaussian observation penalty function.
3AO	As 3A. Fewer data (25 instead of 35).
3AT	As 3A. Observations from higher resolution "truth" (expt A).
3AS	As 3A. Error variances reduced by factor of 10.
3AG	As 3A. Background penalty omitted from J and J'.

LEGENDS FOR FIGURES.

Fig.1. Horizontal correlation of error of background wb used in experiments. Cross-correlations between variables were assumed to be zero. All variables had the same auto-correlation.

Fig.2. Height fields, plotted at time T=0.0. Curves for each experiment are displaced. Each curve has the corresponding curve from the "truth" shown dashed. The observations used are shown as *, with the height of the * showing 6 times the assumed standard deviation of observational error, and the width showing the horizontal gridlength. Graduation marks on the vertical scale are separated by one unit of (non-dimensional) h. Experiments shown are listed in table 1.

Fig.3. As Fig.2 for T=1.4.

Fig.4. As Fig.2 for T=2.8.

Fig.5. As Fig.2 for U at the centre of the grid, plotted against time.

Fig.6. As Fig.2 for wind speed squared at T=0.0.

Fig.7. As Fig.2 for wind speed squared at T=1.4.

Fig.8. As Fig.2 for V at T=0.0, from experiments with zero Coriolis parameter, so that V was a simple tracer.

Fig.9. As Fig.2 for V at T=1.4, from experiments with zero Coriolis parameter, so that V was a simple tracer.

Fig.10. As Fig.2 for V at T=2.8, from experiments with zero Coriolis parameter, so that V was a simple tracer.

Fig.11. Solid line: quadratic (L2) penalty function for a single observation, plotted against the normalized deviation dy[i]/sqrt(O[i]). Dashed line: the equivalent penalty function derived assuming that the observation has a 5% chance of being useless because of a gross error.

Fig.12. As Fig.2 for quality control experiments listed in Table.2. The experiment labels indicated their characteristics: Non-Gaussian error distributions => Q in label. A gross error => G in label. The analysis used as first-guess is indicated at the end of label.

Fig.13. As Fig.3 for quality control experiments listed in Table.2. The experiment labels indicated their characteristics: Non-Gaussian error distributions => Q in label. A gross error => G in label. The analysis used as first-guess is indicated at the end of label.

Fig.14. (a) Reduction factor in total penalty from that of background field, (b) Background penalty, normalized by background penalty of "truth", plotted against iteration, for various descent algorithms. Experiments (listed in Table.3) are: solid line - 1A, long dashes - 1AM, dots - 2A, short dashes - 3A, dot dash - 4A.

Fig.15. As Fig.14, for experiments started from a first guess equal to the "truth".

Fig.16. As Fig.14, for linear, space-only analysis experiments. Experiments (listed in Table.3) are: solid line - 1B, dots - 2B, short dashes - 3B.

Fig.17. As Fig.2, for experiments listed in Table.3, testing approximations.

Fig.18. As Fig.17, for T=1.4.

Fig.19. As Fig.17, for T=2.8.

Fig.20. As Fig.14, for experiments investigating approximations. Experiments (listed in Table.3) are: solid line - 3AS, long dashes - 3AT, dots - 3AG, short dashes - 3A.

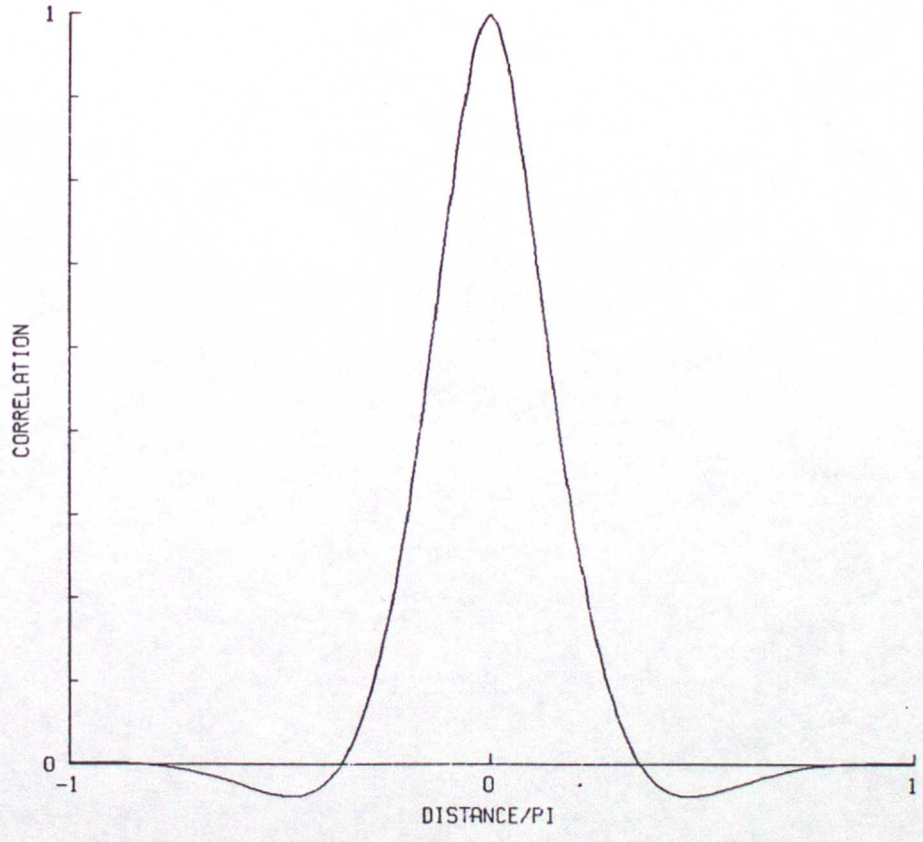


Fig.1. Horizontal correlation of error of background wb used in experiments. Cross-correlations between variables were assumed to be zero. All variables had the same auto-correlation.

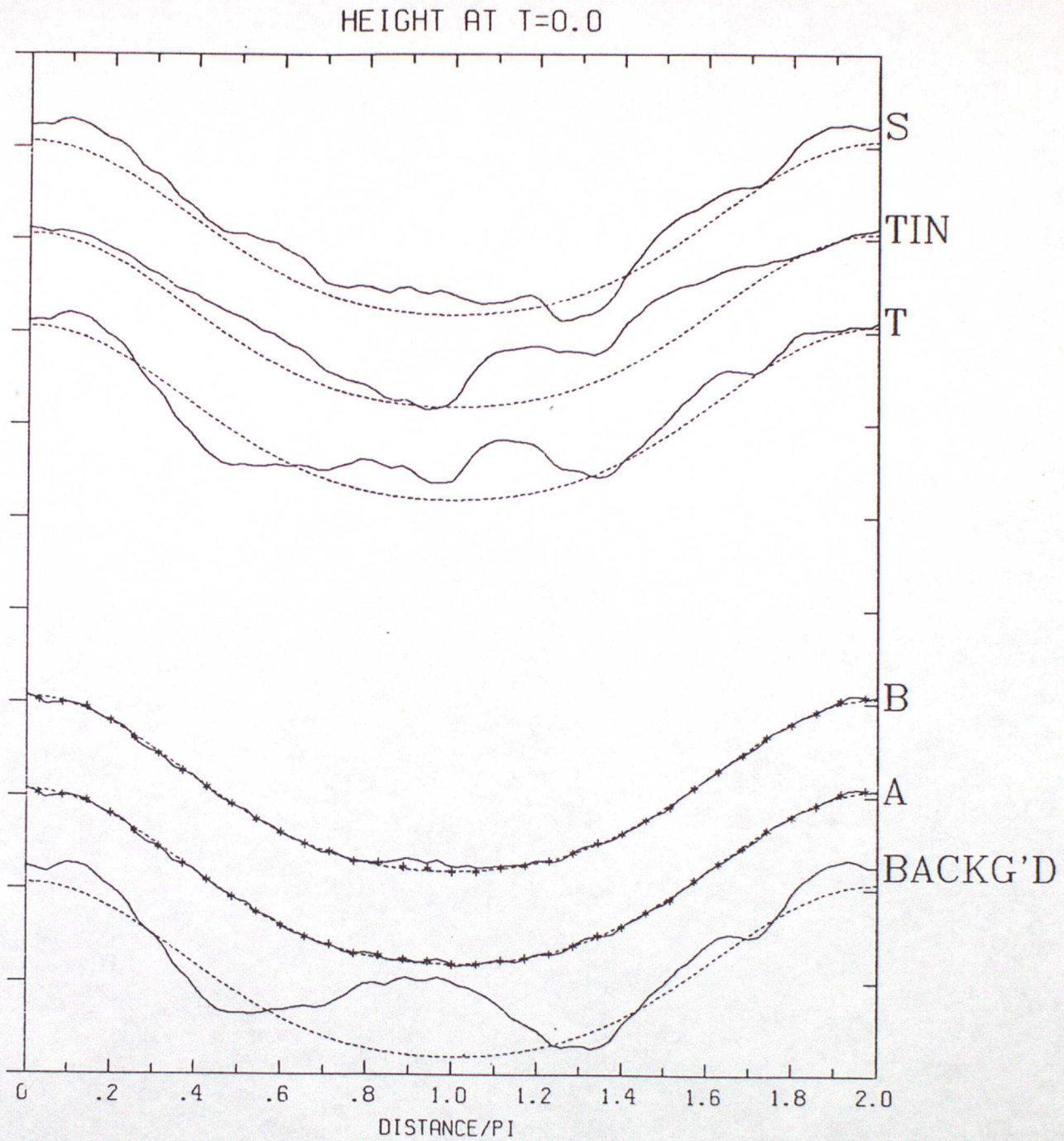


Fig.2. Height fields, plotted at time T=0.0. Curves for each experiment are displaced. Each curve has the corresponding curve from the "truth" shown dotted, and the observations used shown as *, with the height of the * showing 6 times the observational error and the width showing the horizontal gridlength. Graduation marks on the vertical scale are separated by one unit of (non-dimensional) h. Experiments shown are listed in table 1.

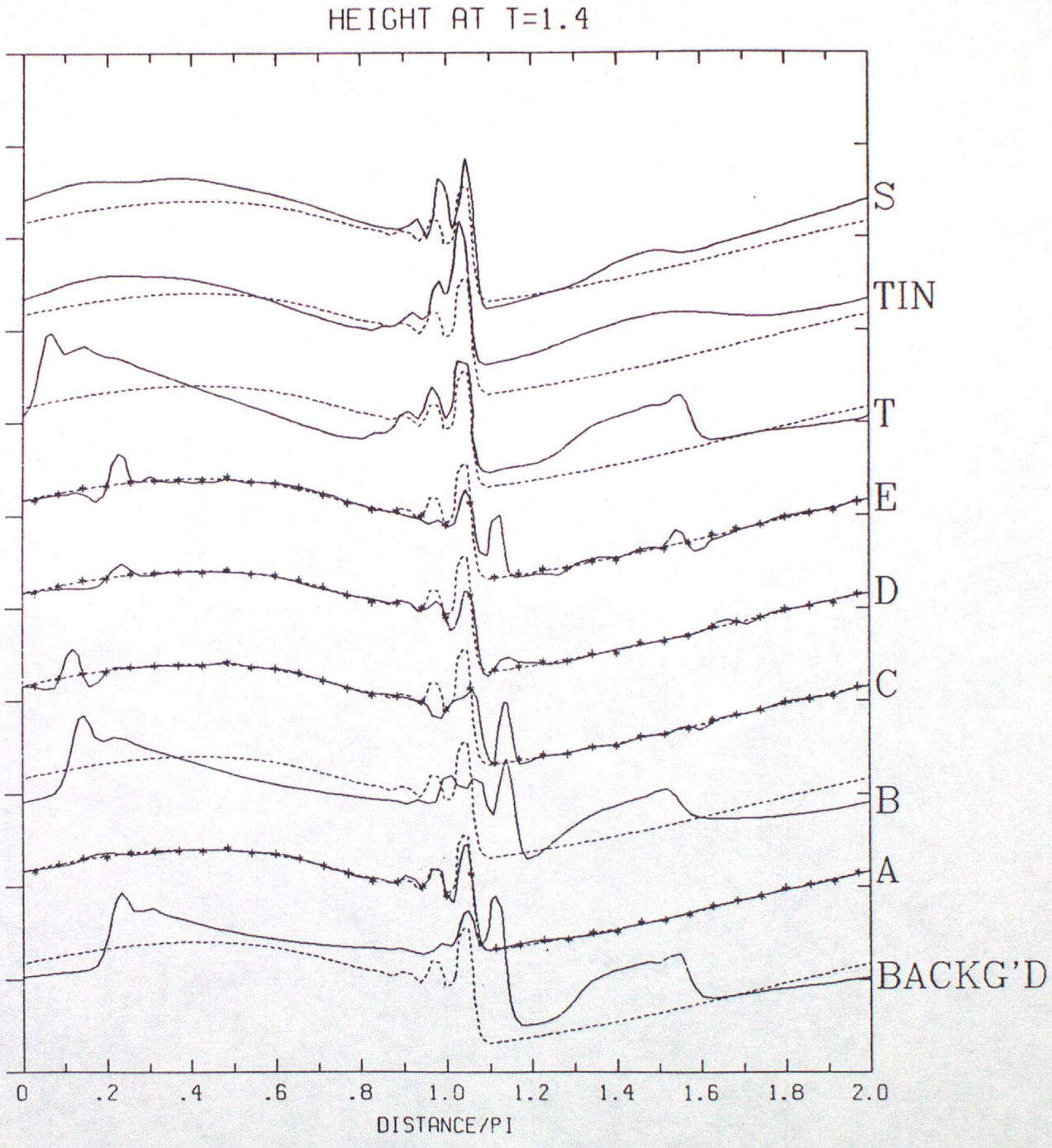


Fig.3. As Fig.2 for T=1.4.

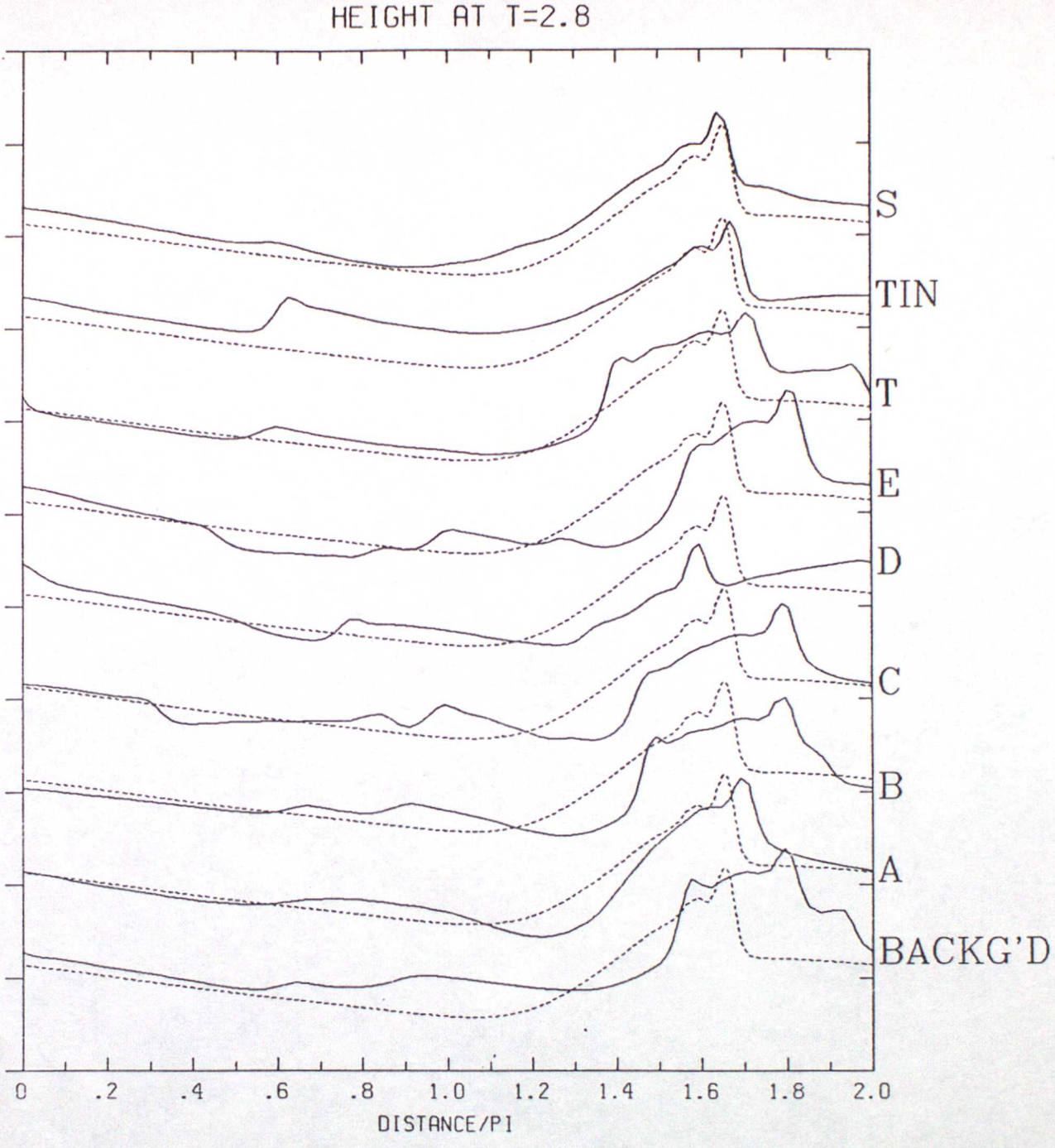


Fig.4. As Fig.2 for T=2.8.

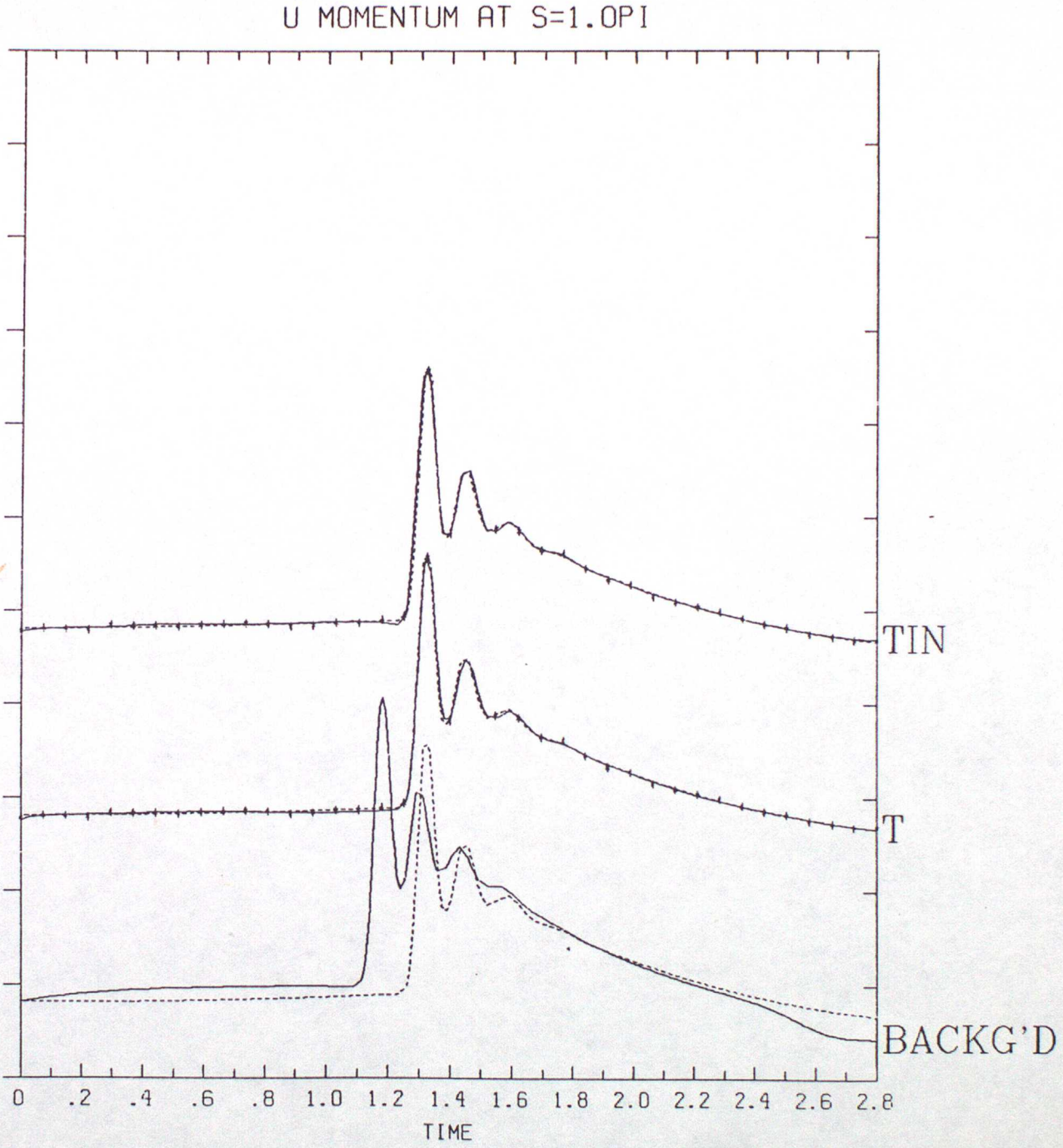


Fig.5. As Fig.2 for U at the centre of the grid, plotted against time.

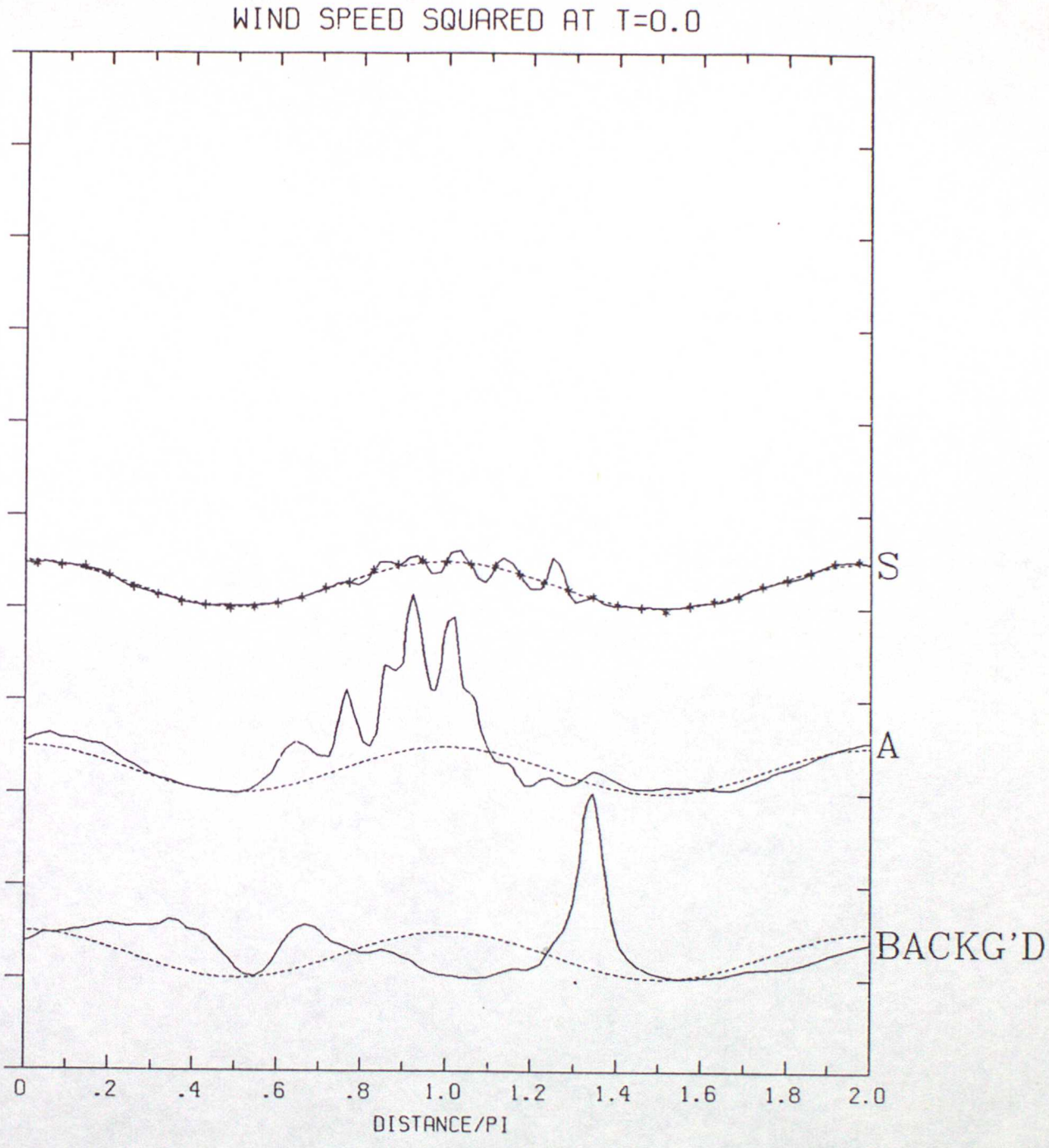


Fig.6. As Fig.2 for wind speed squared at T=0.0.

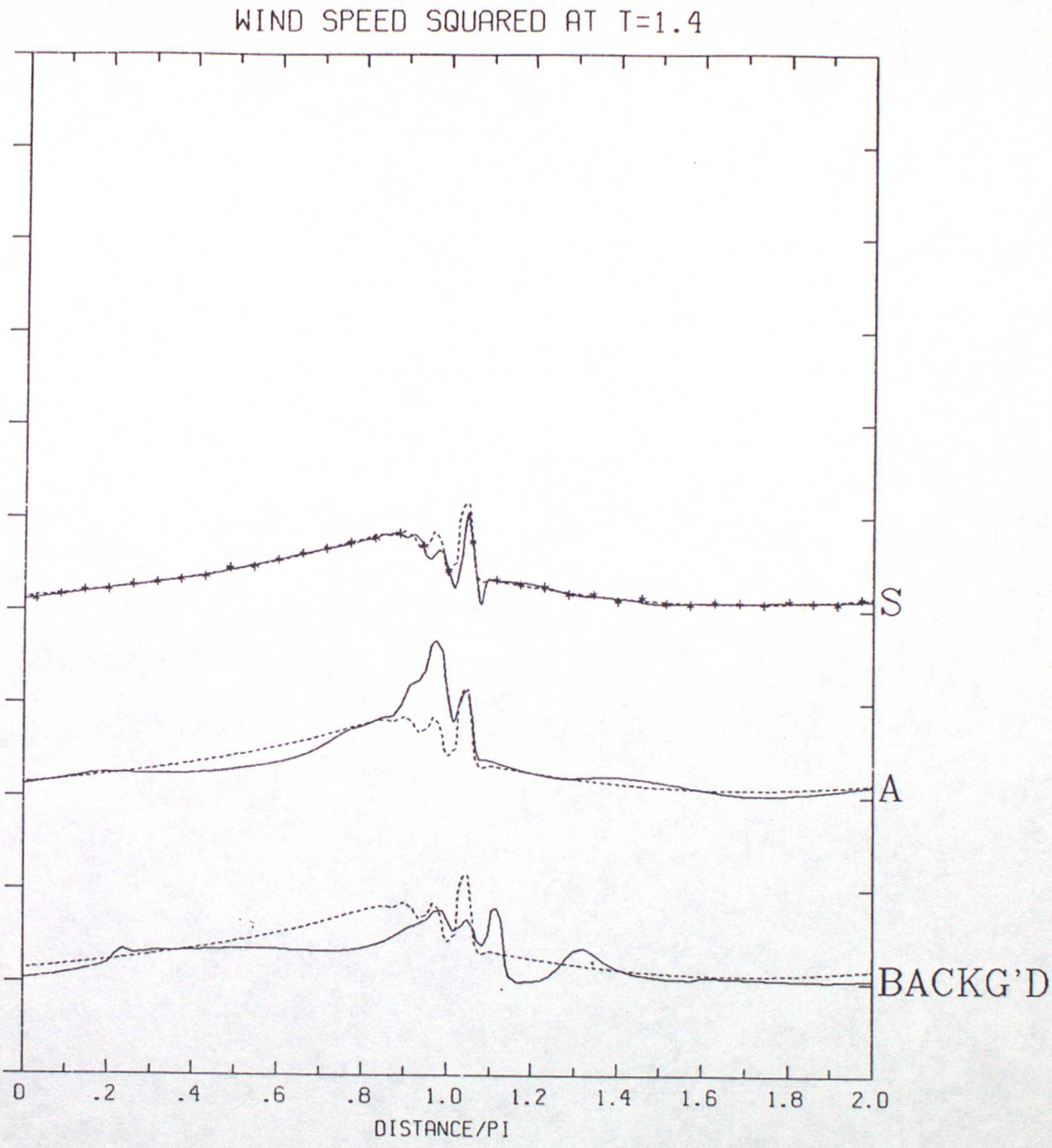


Fig.7. As Fig.2 for wind speed squared at T=1.4.

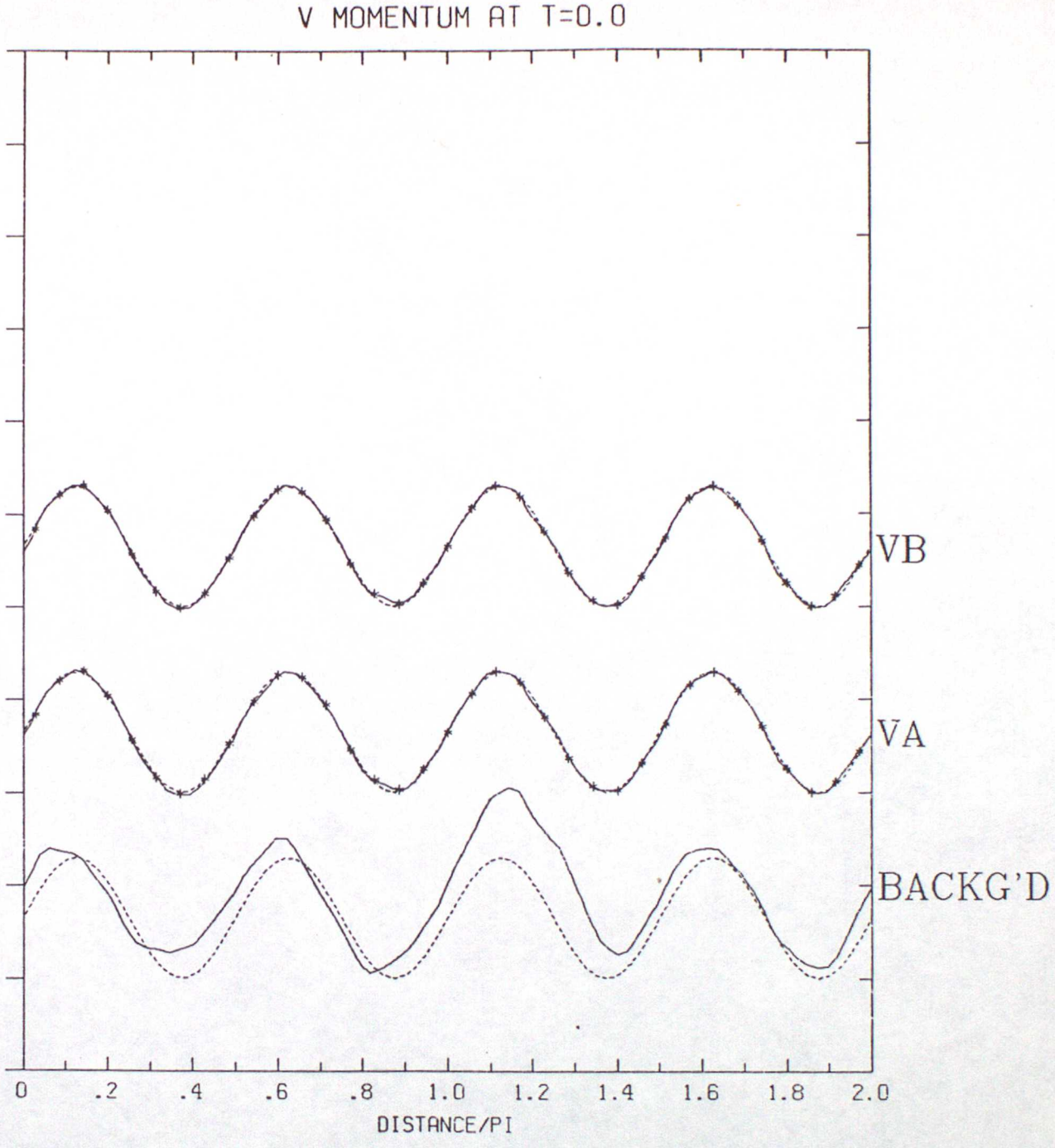


Fig.8. As Fig.2 for V at T=0.0, from experiments with zero Coriolis parameter, so that V was a simple tracer.

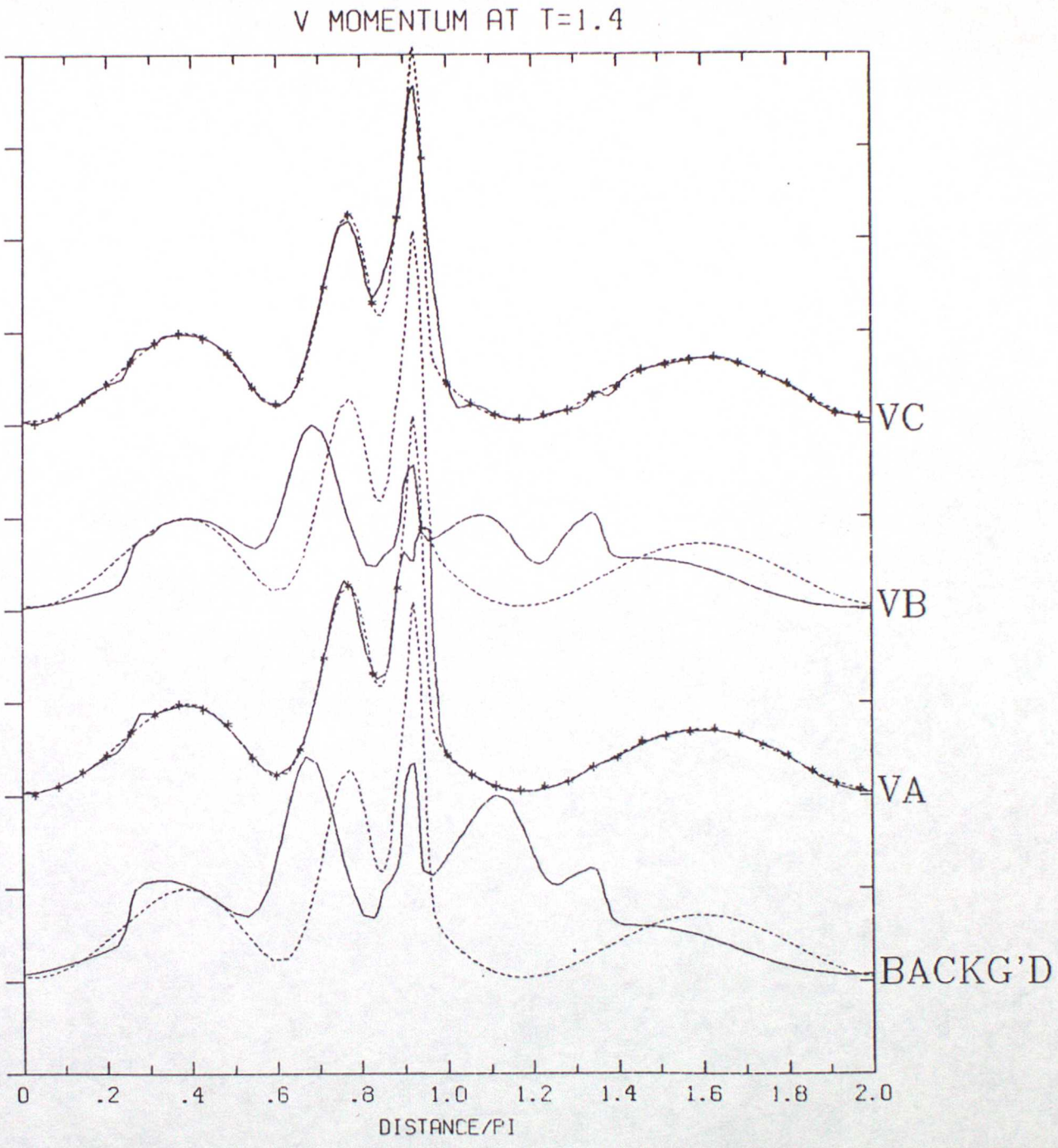


Fig.9. As Fig.2 for V at T=1.4, from experiments with zero Coriolis parameter, so that V was a simple tracer.

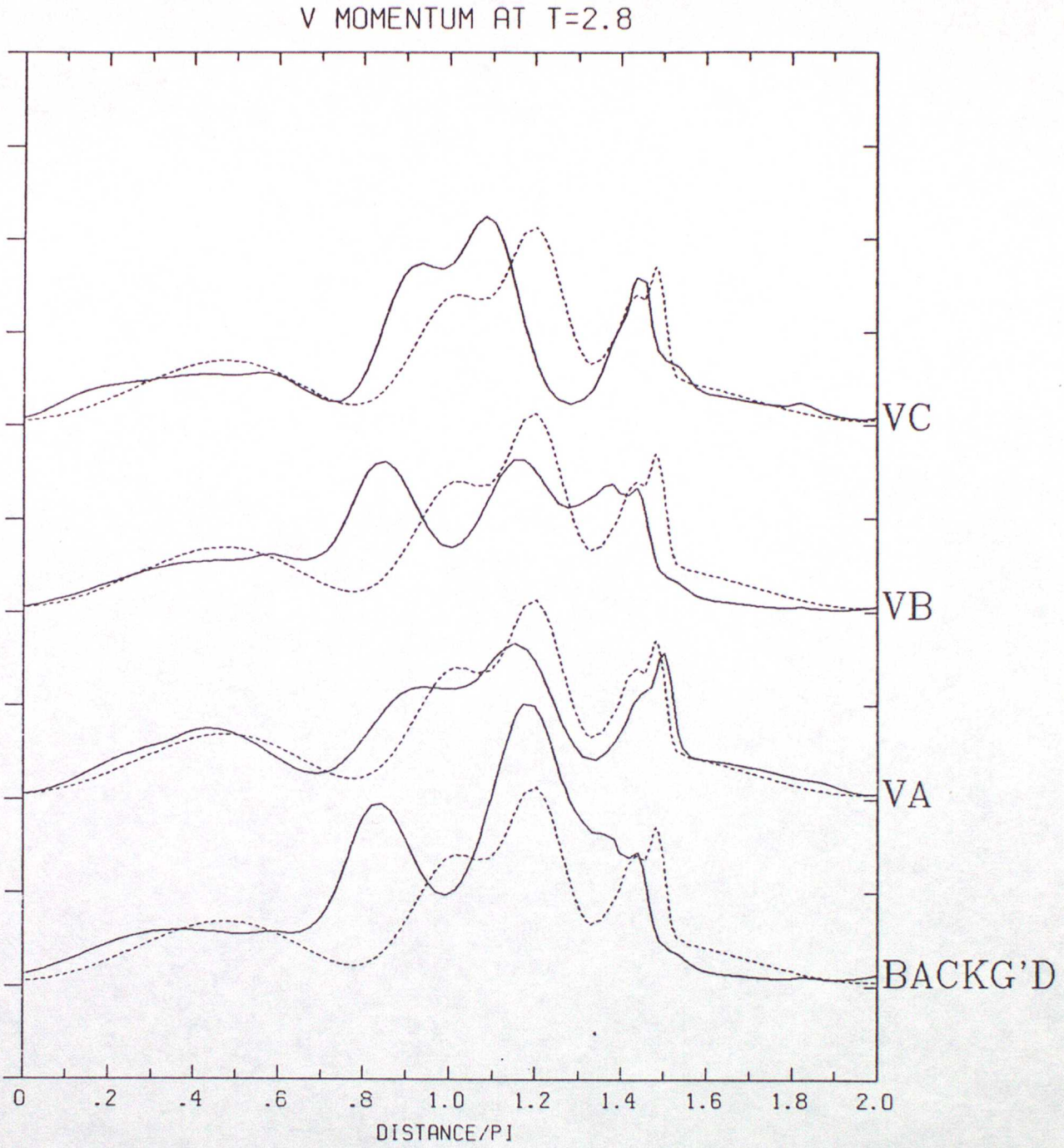


Fig.10. As Fig.2 for V at T=2.8, from experiments with zero Coriolis parameter, so that V was a simple tracer.

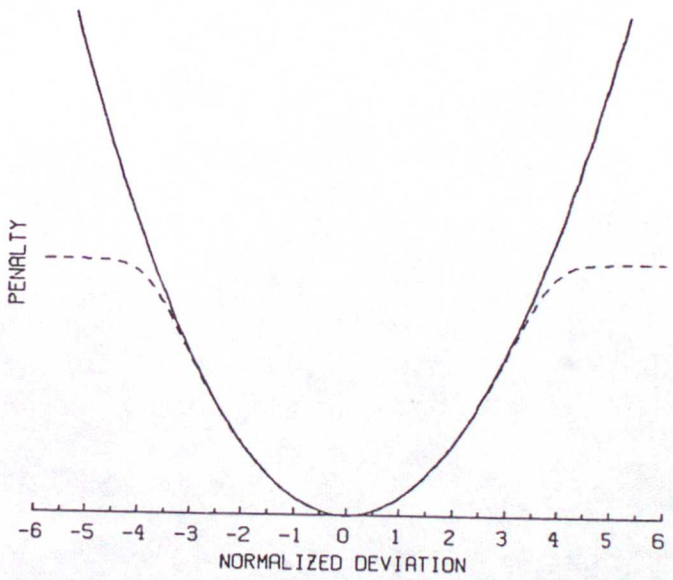


Fig.11. Solid line: quadratic (L2) penalty function for a single observation, plotted against the normalized deviation dy[i]/sqrt(O[i]). Dashed line: the equivalent penalty function derived assuming that the observation has a 5% chance of being useless because of a gross error.

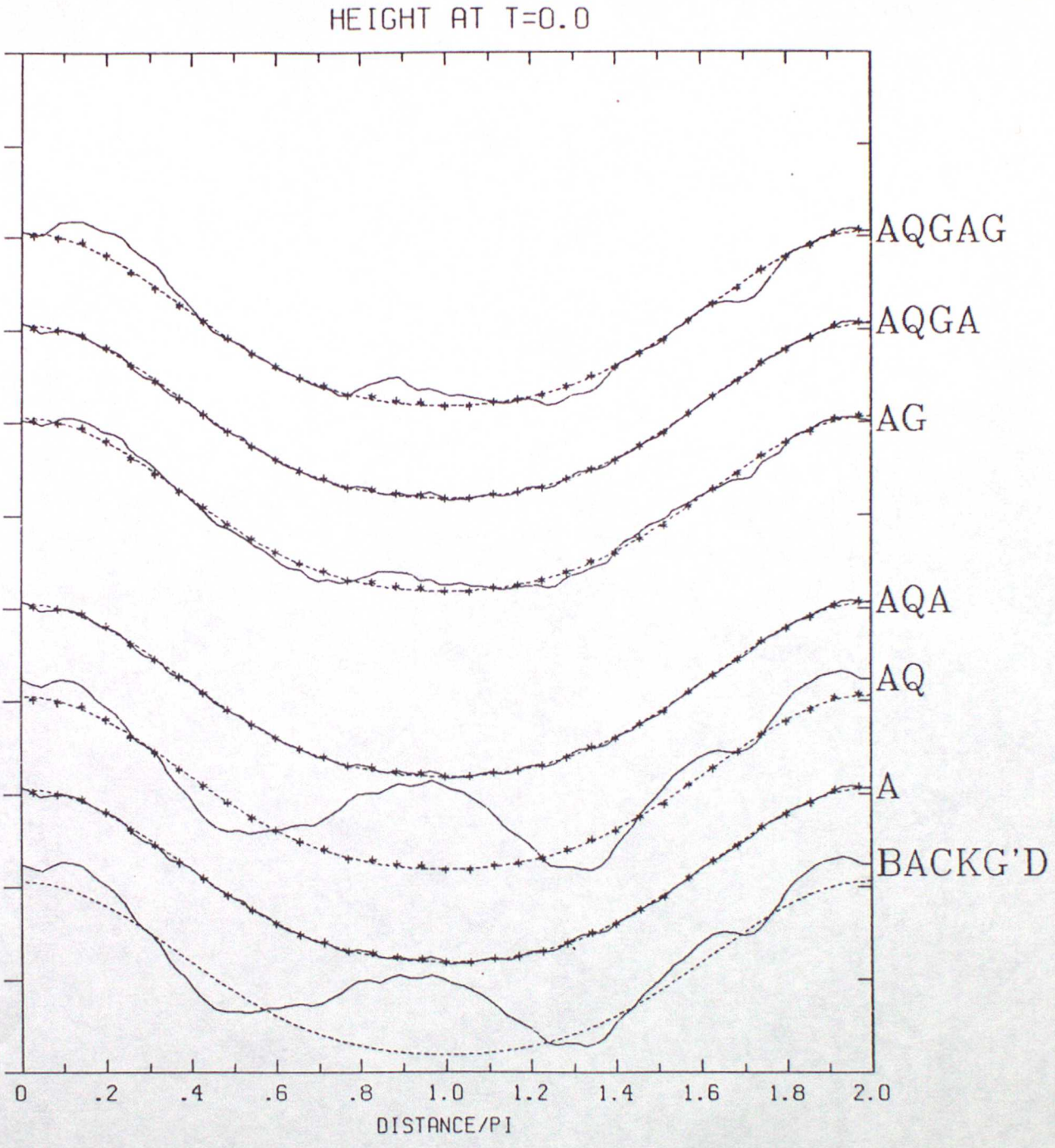


Fig.12. As Fig.2 for quality control experiments listed in Table.2. Non-Gaussian error distributions => Q in label. A gross error => G in label. Analysis used as first guess indicated at end of label.

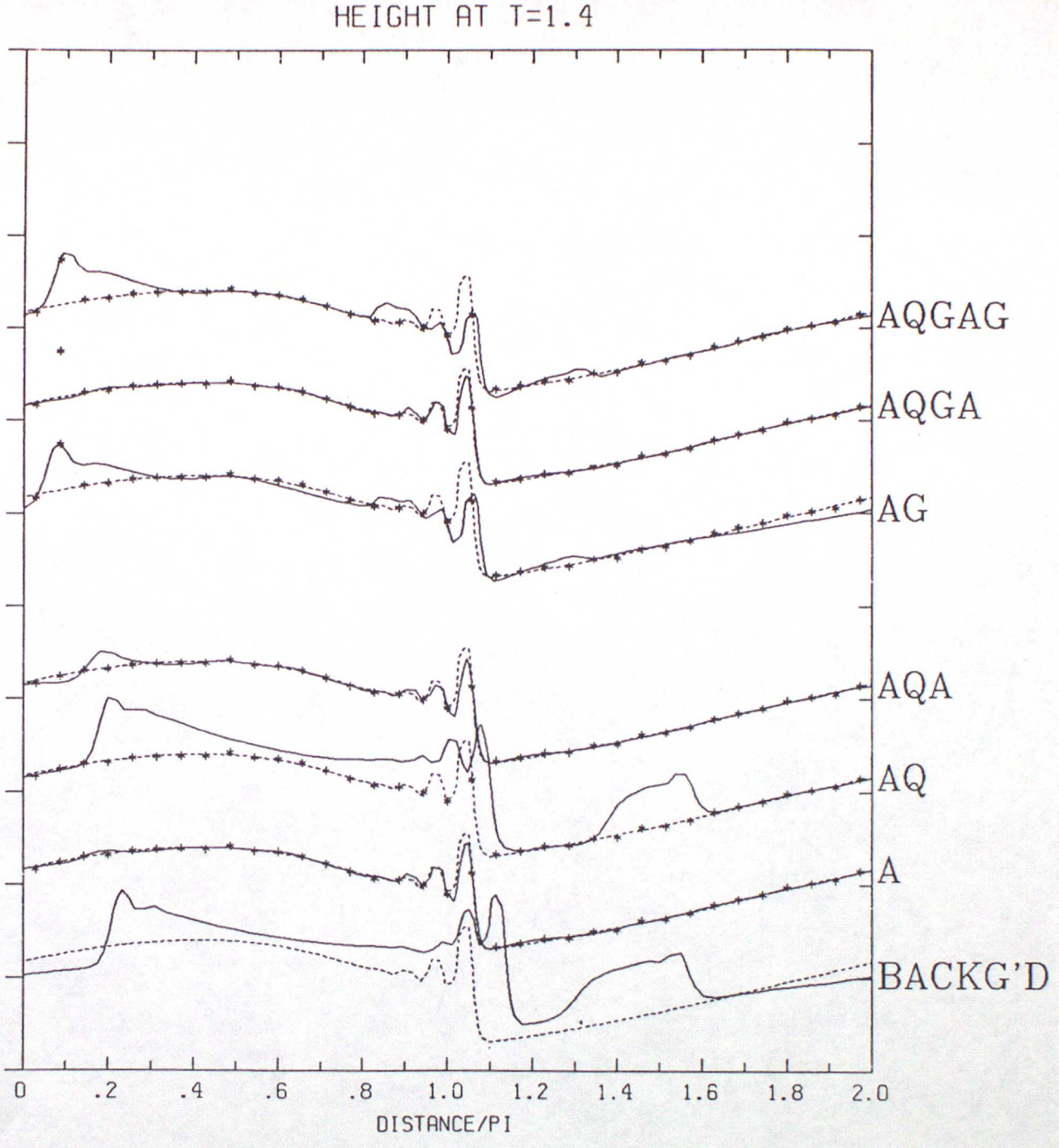
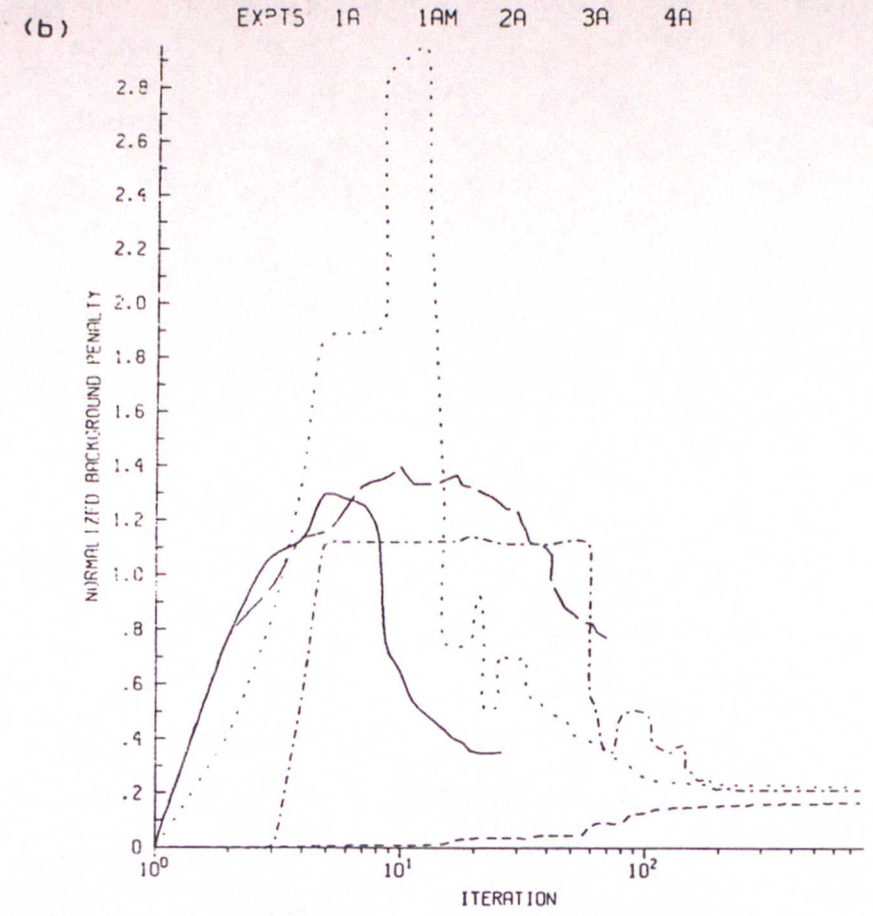


Fig.13. As Fig.3 for quality control experiments listed in Table.2. Non-Gaussian error distributions => Q in label. A gross error => G in label. Analysis used as first guess indicated at end of label.



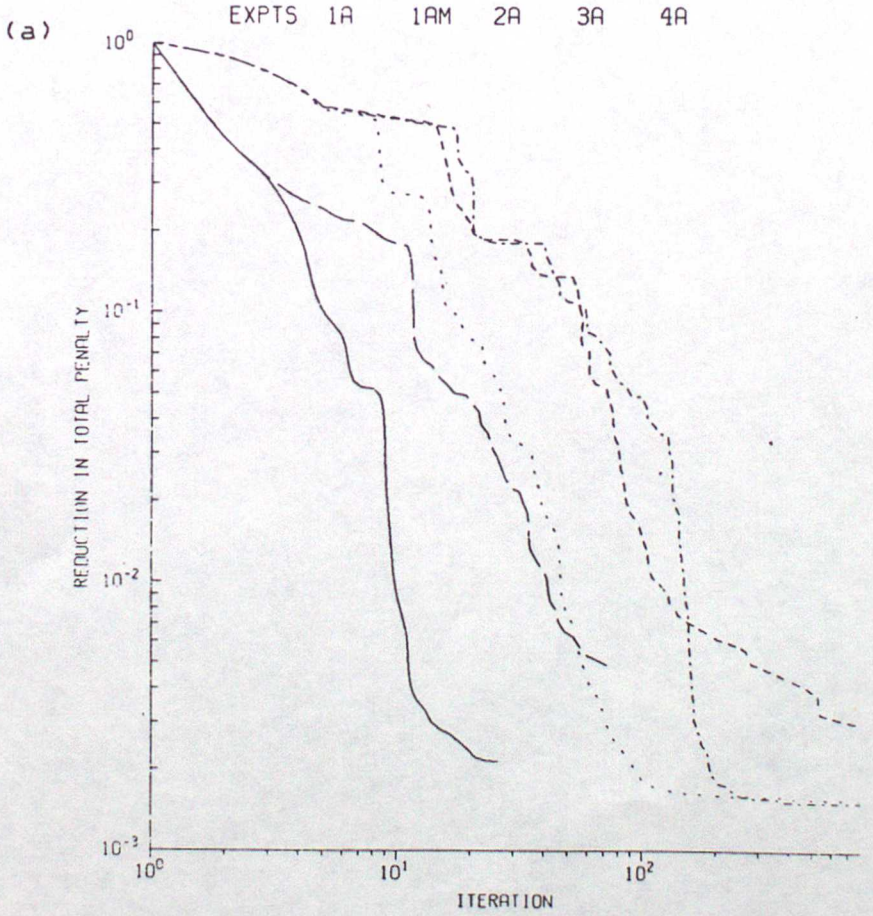
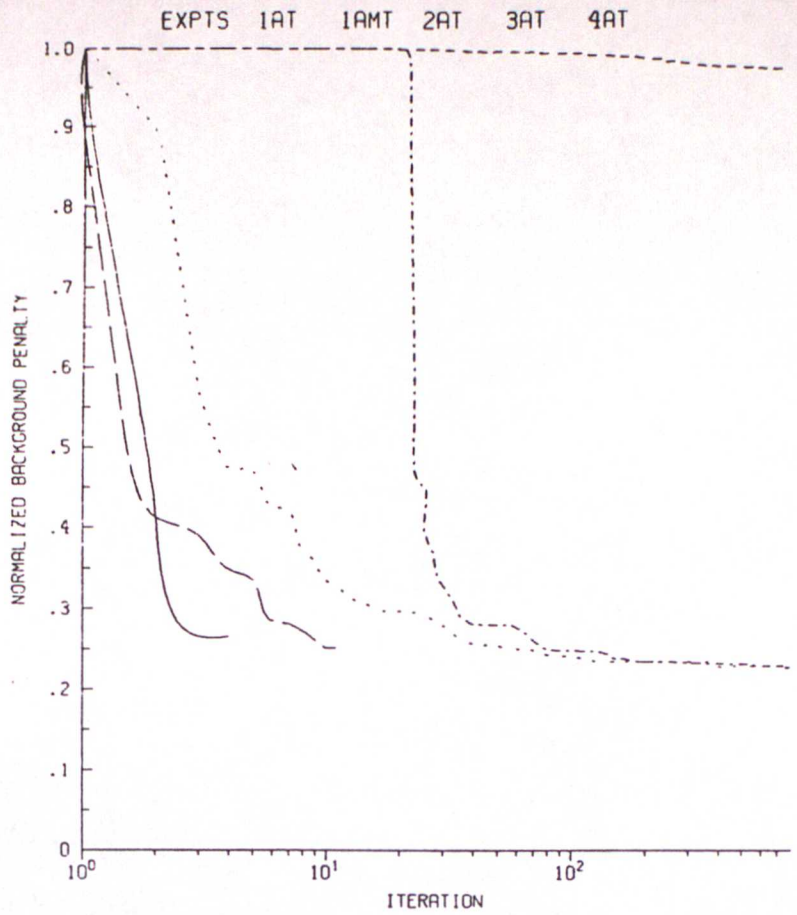


Fig.14. (a) Reduction factor in total penalty from that of background field, (b) Background penalty, normalized by background penalty of "truth", plotted against iteration, for various descent algorithms. Experiments (listed in Table.3) are: solid line - 1A, long dashes - 1AM, dots - 2A, short dashes - 3A, dot dash - 4A.



(a)

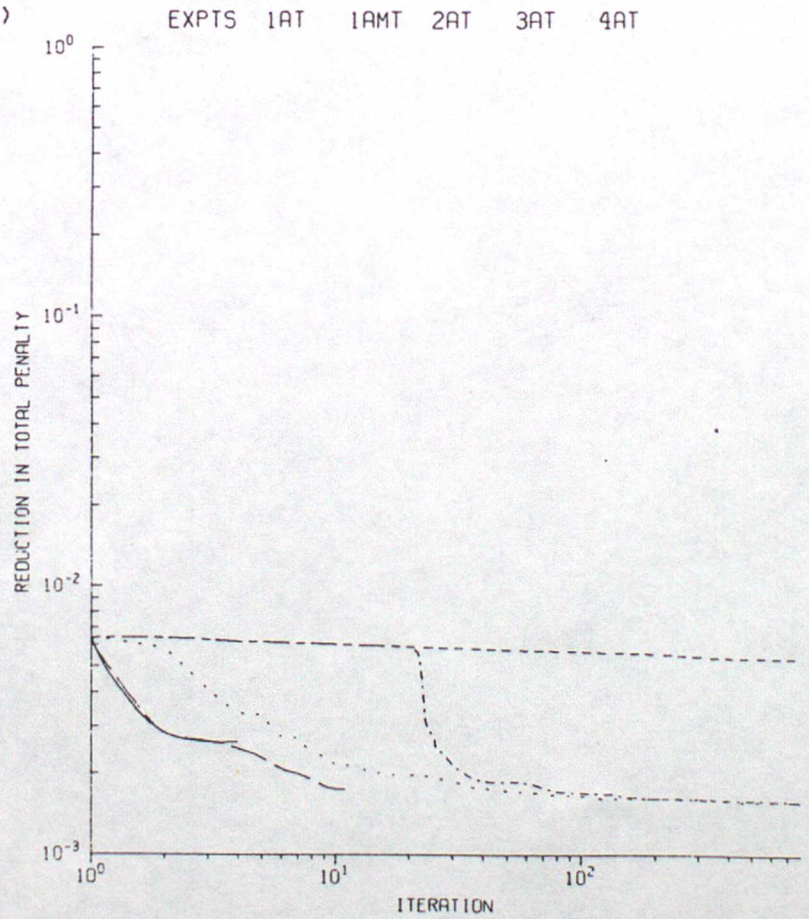
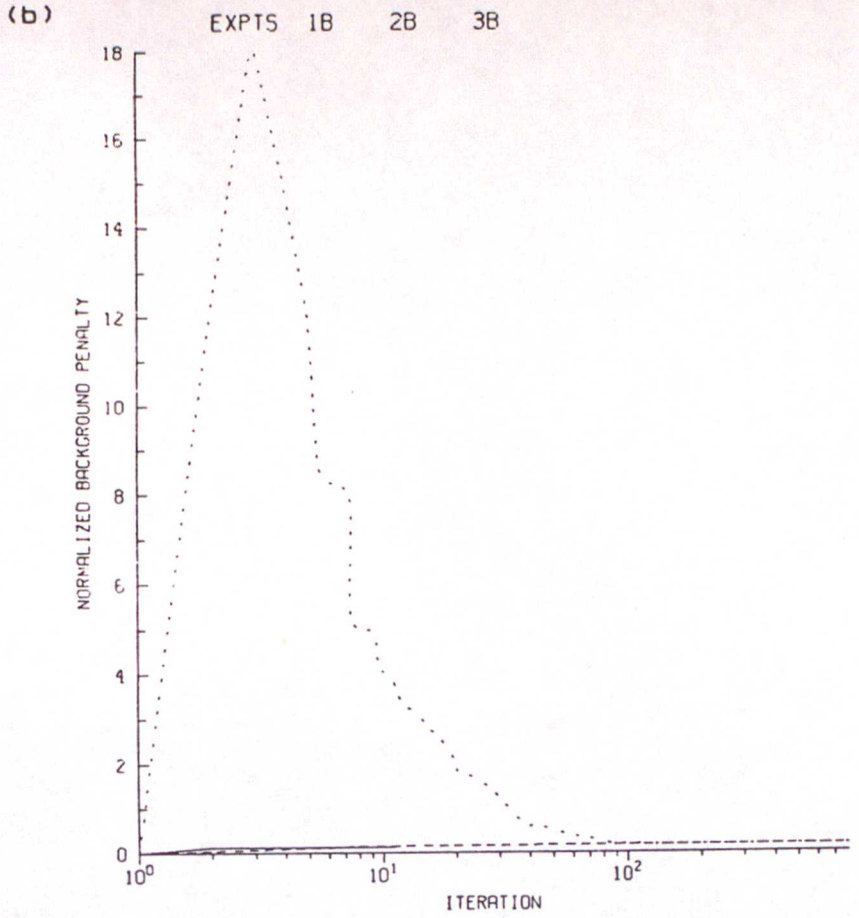


Fig.15. As Fig.14, for experiments started from a first guess equal to the "truth".



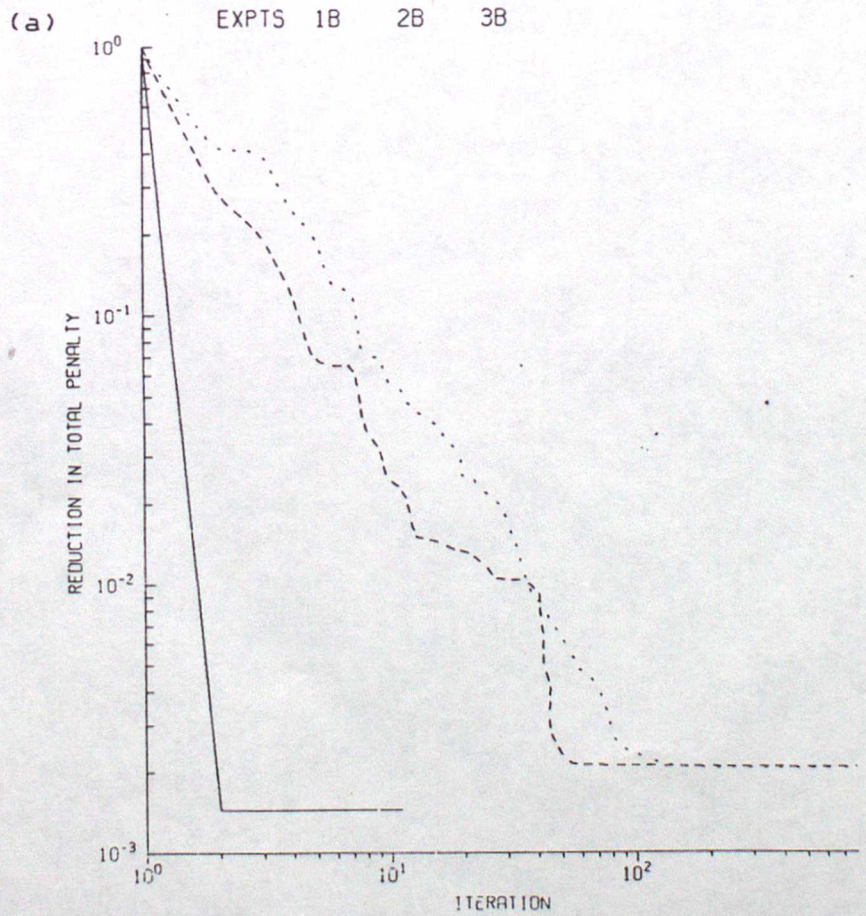


Fig.16. As Fig.14, for linear, space-only analysis experiments. Experiments (listed in Table.3) are: solid line - 1B, dots - 2B, short dashes - 3B.

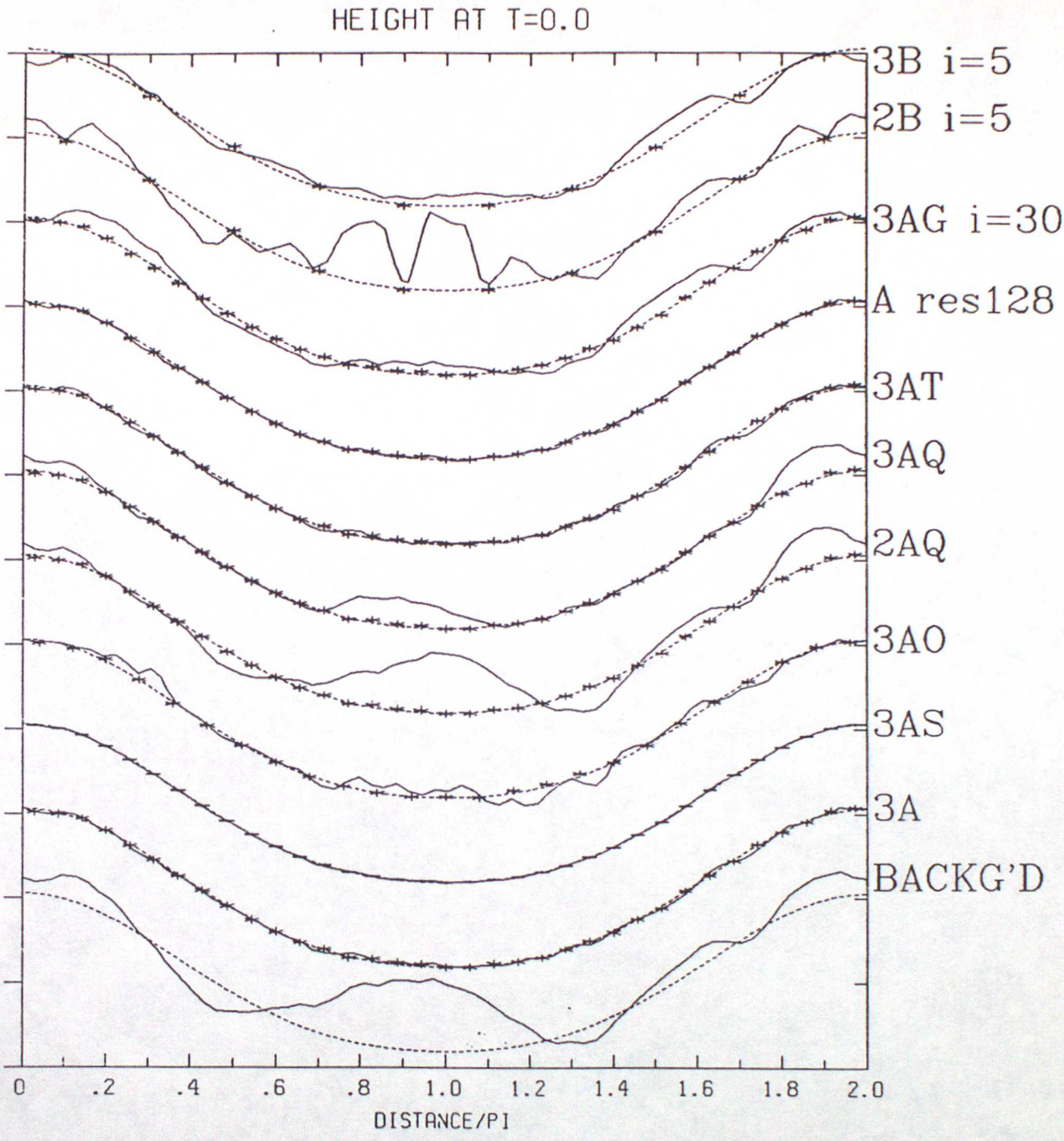


Fig.17. As Fig.2, for experiments listed in Table.3, testing approximations.

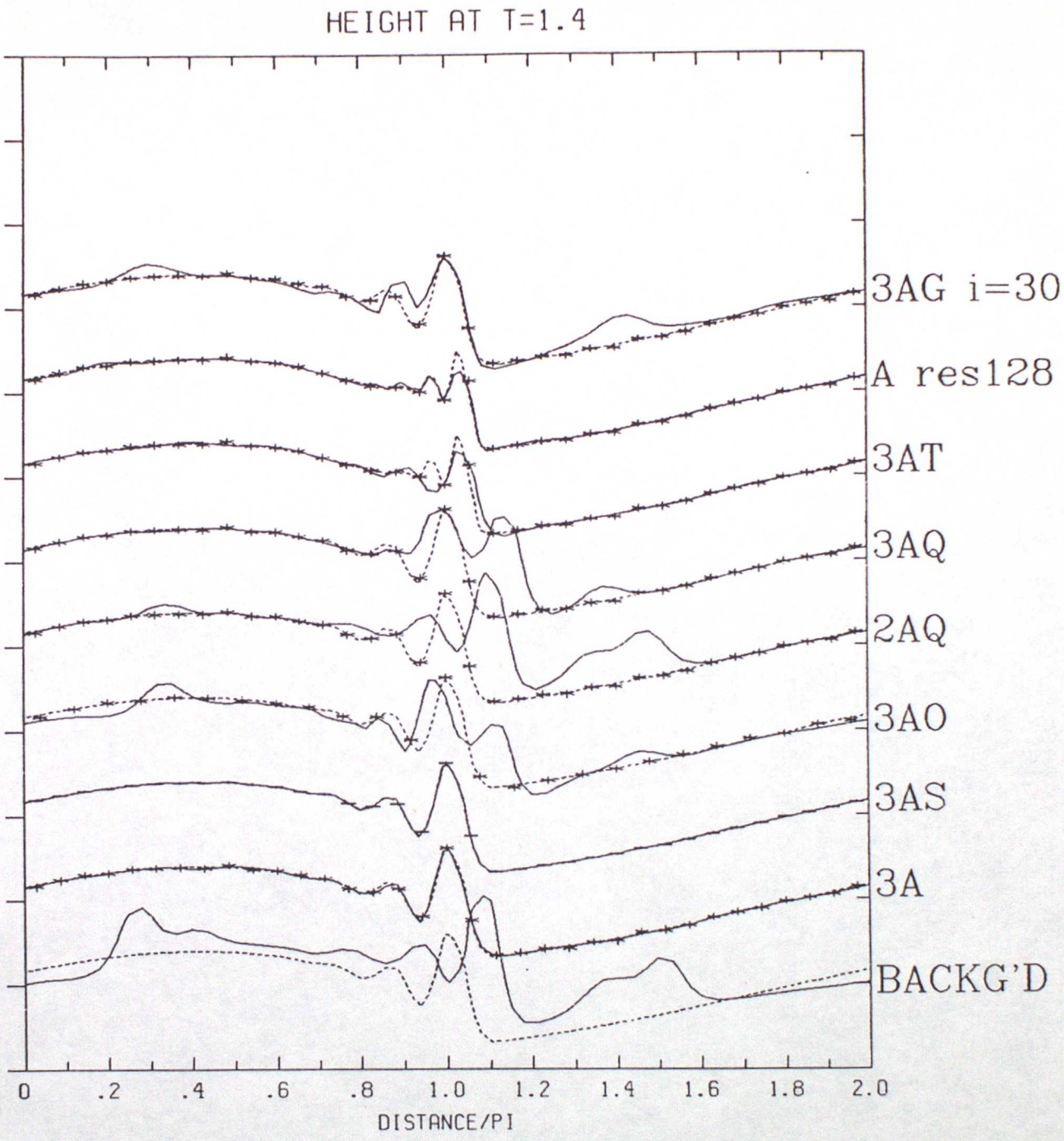


Fig.18. As Fig.17, for T=1.4.

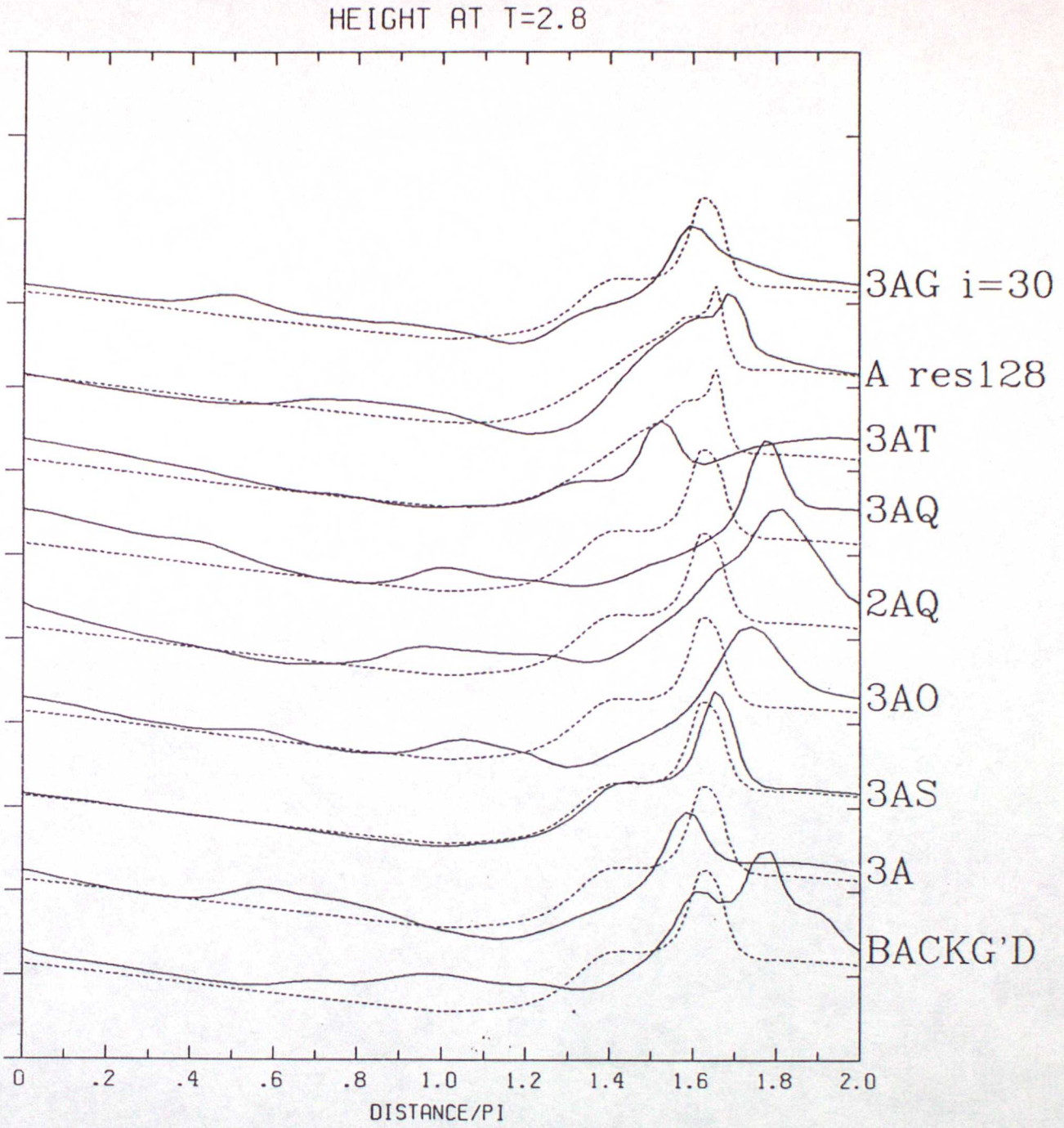
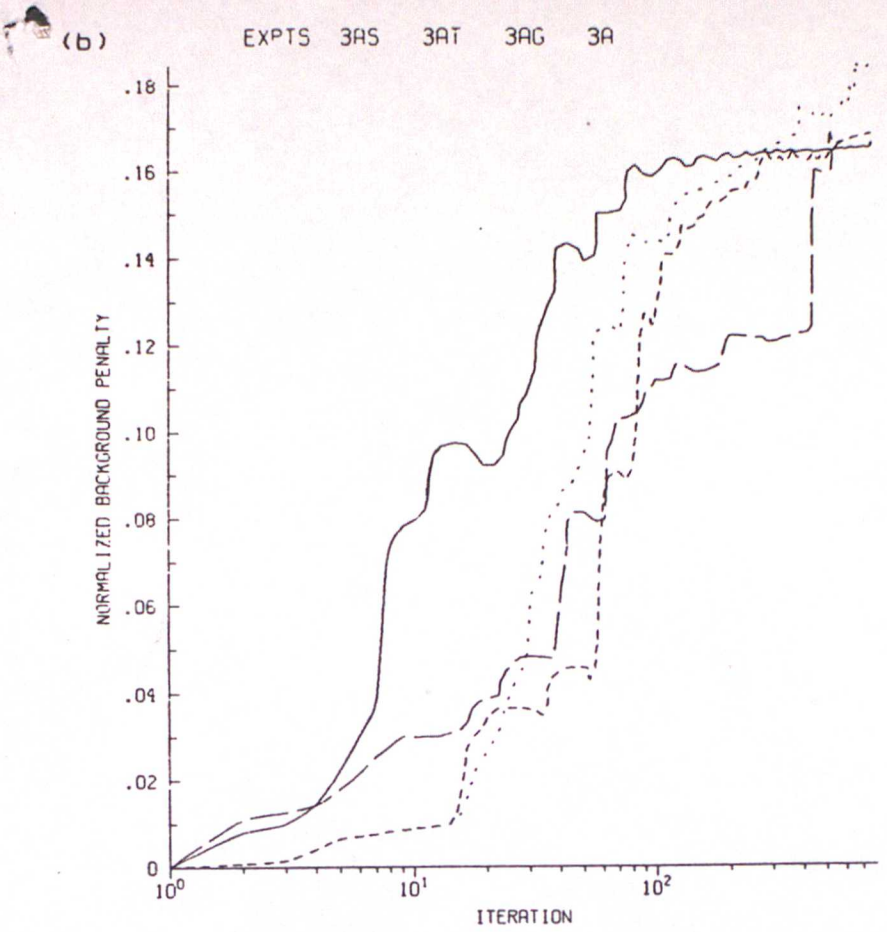


Fig.19. As Fig.17, for T=2.8.



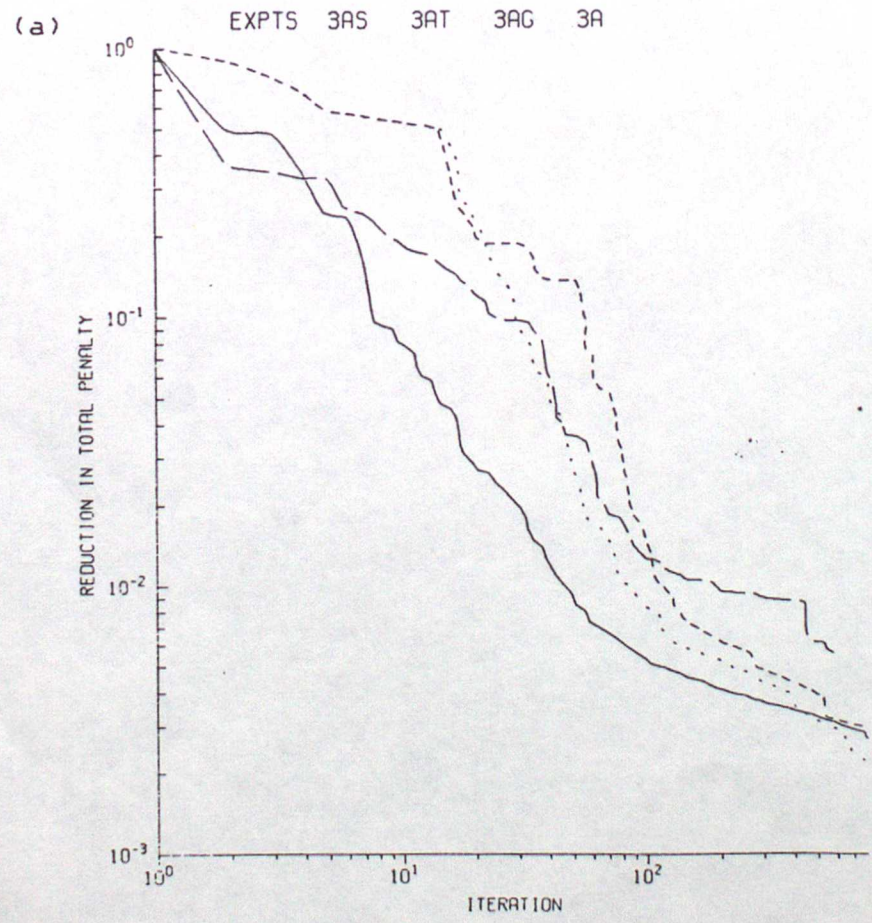


Fig.20. As Fig.14, for experiments investigating approximations. Experiments (listed in Table.3) are: solid line - 3AS, long dashes - 3AT, dots - 3AG, short dashes - 3A.