TECHNICAL MEMORANDUM

150

Design of a variational analysis: organisation and main scientific points. Computation of the distance to the observations

J. Pailleux

Research Department

February 1989

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Abstract

A comprehensive and practical approach of the 4 dimensional assimilation problem is presented in the ECMWF seminar proceedings on data assimilation and use of satellite data (September 1988), and also in Talagrand and Courtier (1987), Courtier and Talagrand (1987) and Lorenc (1988). This approach uses variational techniques which minimizes the distance between the forecast model and the observations: it is called "4DVAR" (4 dimensional variational) assimilation. The distance between the model and the observations is called "cost function" in this paper (it is sometimes called "misfit" function or "penalty" function). The technique relies on the notion of an adjoint operator which provides a very convenient tool for computing the gradient of the cost function with respect to the variables of the forecast model X (X is the "control variable" of the variational problem). The gradient of the cost function is required by all the standard minimization algorithms in addition to the control variable and the value of the cost function.

The present document is the scientific design of the part of the variational analysis dealing with the distance of the model to the observations. It involves only 3-D aspects as all the aspects related to the time evolution are treated through the forecast model and its adjoint. After having set up the ECMWF environment of the variational analysis (section 1), and having considered the general scientific aspects (section 2), the present paper examines the scientific aspects of the processing we need to apply to each observing system (sections 3 to 10).

1. INTRODUCTION: CONTEXT OF THE DEVELOPMENT OF THE VARIATIONAL ANALYSIS

An "in-core system" is being developed at ECMWF: it contains an in-core forecast model, the corresponding tangent linear model and its adjoint. It contains also the general architecture needed to run 4-D assimilation experiments: minimization scheme, plus all the facilities to "plug" a variational analysis.

From the purely scientific point of view, the variational analysis consists of computing a cost function J(X) and its gradient with respect to X (X is the vector containing all the model variables at the initial time of the assimilation period.) J(X) and Grad J are then passed to the minimization scheme.

The total procedure is repeated several times until an appropriate convergence is reached.

$J = J_o + J_g + J_c$ with:

- J_o : distance of the model X to the different observations (X now taken at any time of the assimilation period).
- J_g : distance of X to a first guess X_g.
- J_c : cost functions describing some physical constraints on the fields. Let us note that the use of J_c is not the only technical way to insure some physical constraints.

A variational analysis reproducing almost exactly the present analysis would contain the geostrophic constraint on the increments as J_c, $J_o = (HX-d)^t\ O^{-1}\ (HX-d)$ and $J_g = (X-X_g)^t\ P^{-1}\ (X-X_g)$, O being the covariance matrix of the obs error, P the covariance matrix of the forecast error for all the model variables, H the mathematical operator going from the model variables to the vector of observed data d. (See RD memos R2327/1917 by Courtier and Pailleux, and R2327/1965 by Pailleux. See also the ECMWF proceedings of the September 1988 seminar).

The different terms of the cost functions can be designed and computed independently.

For a flexible and general design of a 4-D variational analysis, we need comprehensive specifications for all the input to the analysis, which are:

- All the observations (vector d);

- A statistical set describing all the observation errors (matrix O);

- The first-guess (vector X_g);

- A statistical set describing all the forecast errors (matrix P);

- An estimate of the covariance matrix of the analysis error for the model variable (i.e. the "Hessian" to use in the minimization procedure).

The system which is designed in the present document is to be used to run comprehensive research experiments on the present computers, and, assuming this research is successful, to run an operational 3-D variational analysis on the next super-computer (around 1991). In this context, it is assumed that the variational analysis will be run on top of a traditional Optimum Interpolation (OI) analysis. The OI analysis will be used for two main purposes: - to provide an initial point for the minimization scheme, which is close to the final solution; - to perform a quality control on the data.

For both scientific and practical reasons it might be more convenient to run this preliminary OI analysis without satellite data. This would at least provide the framework for a better quality control of satellite data. Assuming most of the quality control is performed by the preliminary OI analysis, an extra level of quality control will be developed in the context of the variational analysis: - because it is necessary if in the variational analysis some data are used which do not enter the OI analysis; - because the 4-D variational context provides a framework to perform extra quality control checks which are more difficult to set up in the traditional OI. This last point is discussed in section 2.6.

2. GENERAL STRATEGY

2.1 How to split the cost function J_o

J_o can be written $E^t O^{-1} E$ where O is now the correlation matrix of the observed data, and E is the vector of the departures "observed - HX" normalized by the observation error standard deviation. Careful: we are used to normalizing the departure by the forecast error, but in this case we must normalize by the observation error. The reason for this normalization is that it is easier in practice to work with a correlation matrix O than with a covariance.

As there is no correlation between obs errors of different types, the different contributions of the different obs types will be computed independently: $J_o = J_{synop} + J_{rs} + J_{pilot} + J_{satem} + \ldots$. For diagnostic purposes they will have to be printed out (or even stored for the different iterations of the minimization procedure, say, the first and the last one as an example). The general strategy will be also to split the contribution of one observation type into as many different sub-contributions as we are interested in. This will be useful in order to understand how the variational analysis is working and also to detect some data problems. Examples: compute separately the contributions of land SYNOPs and SHIPs, separate also the contributions of the different geostationary satellites for SATOBs...

2.2 The input to the computation of J_o

The computation of J_o has to be made from two different sources of information:

- The observations which have to be stored before the variational analysis on an ad hoc file. The file should be in a transportable format, it should contain the comprehensive information, including the information on the quality control and event flags coming from the optimum interpolation analysis. The observations should be sorted observation type per observation type (see 1.1).

- The model variables in grid points as they are available in central memory in the in-core system.

In addition we need a third source of information: the covariance matrix of observation errors for all types of observations, keeping in mind that the observation error should also include the error of the operator H, that is the representativeness error.

2.3 Computation of HX

The operator H which must be applied to the model vector X in order to get the data at the observation points, is a "post-processing" operator which has to be studied for each observation type. For a spectral model, X is in the spectral space. However, we can assume that X contains the model variables on a latitude/longitude grid, as all the spectral transforms are already designed in the context of the in-core system.

For each observation type, H contains a bilinear interpolation in the horizontal to the observation point, from the four nearest grid-points. In the vertical, depending on the observation type, H contains operators such as:

- ordinary vertical interpolation (e.g.: interpolation of the model wind profile to the level of an observed wind);

- integrations involving both T and Q (e.g.: computation of a geopotential height from model variables, for radiosondes).

- full radiative transfer computations for radiance data.

Let us call H_v the vertical part of the operator H. H_v includes non-linear operators which are much more complicated than an ordinary vertical interpolation. Then the first processing to apply to X should be the bilinear horizontal interpolation, in order to get the COMPREHENSIVE SET OF MODEL VARIABLES AT THE OBSERVATION POINT X^{mvo}. Then all the remaining parts of HX will be computed from X^{mvo} (and from the observations themselves.)

Very often there is no horizontal correlation between the errors of two different observations. In this case the contribution of each observation to J_o can be evaluated very simply as soon as HX has been computed:

CONTRIBUTION = $E^t \, O^{-1} \, E$ with E equal to the departures d-HX (= $d-H_vX^{mvo}$) normalized by the observation error standard deviation.

2.4 Computation of the gradient of J_o

The gradient of each contribution to J_o with respect to X can be computed through the following steps:

- computation of the gradient with respect to E:

  $$Grad_{(E)} J = 2 \, O^{-1} \, E$$

- computation of the gradient with respect to the vector H_vX^{mvo}:
  As E is equal to the vector d - H_vX^{mvo} normalised by the observation error, the gradient with respect to H_vX^{mvo} is obtained by dividing each component of $Grad_{(E)}J$ by the observation error, and by changing the sign.

- computation of the gradient with respect to the vector X^{mvo}:

  $\text{Grad}_{(X^{mvo})} J = H_V^*(\text{ Grad}_{(H_V X^{mvo})} J)$ where H_V^* is the adjoint of the vertical postprocessing operator H_V.

- multiply the gradient with respect to X^{mvo} by the ad hoc coefficients of the bilinear interpolation in order to get the contribution to the gradient at the four nearest grid-points.

- add these contributions to the general dual variable carried by the in-core system.

Depending on the data type, H_V might be applied either through a simple analytical computation or by coding the subroutine which is the formal adjoint of the routine H_V. For observations which are not correlated in the horizontal, the computation of the gradient can be done independently for each observation, like the computation of the cost function.

2.5 Different steps to envisage in the design

For each observation type the design and coding strategy has to envisage two steps:

a) STEP 1: for benchmark and validation purposes, reproduce with the variational scheme something which is as close as possible to the operational OI analysis, which means using the same pieces of data from the observation as in the present analysis (most of the time).

b) STEP 2: try to use all the information which is available in the observation type. If the same information can be used in two different forms, try to use the one which is closer to the observed quantity. Examples of observations available in different forms: - in radiosonde observations, the geopotential height is redundant with the temperature and humidity profiles; - all the wind data can be handled either by using the components, or by using speed and direction...

2.6 Quality control of observations

Before the variational analysis a traditional OI analysis is expected to be run for at least two reasons: to provide an initial solution to the minimization algorithm, and also to perform the quality control. All the information about the OI quality control has to be passed to the variational analysis (flags, events, departures to the first-guess...). Then the variational analysis can be done without any specific quality control, and all the first research experiments are expected to be run in this way.

However, an extra quality control procedure has to be planned in the variational analysis because:

- We might want to use extra data which are not used in the operational OI system (e.g.: temperature data from parts B of radiosondes);

- We might want to take advantage of the variational context to improve the OI quality control, by using a sequence of model values at the data points for different steps of the minimization algorithm. Possibly one or two iterations of the minimization procedure could be run before the main analysis with the unique goal of checking the data. This would be especially interesting in the 4D context.

The following ideas will have to be tried at some stage through assimilation experiments:

a) IMPROVED TIME CONTINUITY CHECK.
For stations or platforms reporting with a high time frequency, the differences "observation - model" can be examined before the minimization (step 0) and after 1 or 2 steps of the minimization procedure. The time evolution of these differences can be examined as well: performed at step 0 it is nothing but a classical time continuity check which is expected to be helpful for quality control of SYNOPs or buoys in cases such as rapid deepening of a low. The differences "observation - model" compared at steps 0,1,2... for a particular observation, are likely to detect some wrong observations, as the minimization scheme will find it difficult to find a model trajectory which fits the wrong observation as well as the other sources of information (model dynamics, other observations of the assimilation period in the area of the wrong observation...)

b) INTERCOMPARISON OF SATELLITE DATA ON A GIVEN AREA.
Having split the globe in different geographical areas (or different air-masses), the distance of the model to each independent subset of satellite data can be computed over each area, allowing some intercomparisons. Examples: comparison of the distance to the model over a polar cap of:

- NOAA10 clear soundings,
- NOAA10 non-clear soundings,
- NOAA11 clear soundings,
- NOAA11 non-clear soundings,
- each DMSP satellite.

One bad satellite subset is then likely to be detected through the intercomparison of the different distances as it will show a bigger distance than the other subsets. This kind of "regional subset check" can be applied even in the context of the current OI analysis; however a 4D assimilation provides the possibility to improve the satellite intercomparisons, especially if the quality control version of the 4D variational system is run on a 12h period, as the data coverage of each polar orbiting satellite is global on such a 12h period.

All the quality control information of the preliminary OI has to be kept in order to leave open all the possibilities in the variational quality control step (same basic idea as the one suggested by Gandin (1988)). As an example, an observation rejected by the OI must still have the possibility to be used in the variational analysis.

3. COMPUTATION OF J_o FOR SYNOPS

For each SYNOP observation, the contribution to the cost function and its gradient can be computed separately for the following parameters, as the observation errors associated to these parameters can generally be assumed to be uncorrelated:

- surface pressure (or mean sea level pressure);
- 10m wind (not used for the moment from land SYNOPs over most areas);
- 2m temperature (not used in the present mass/wind analysis);
- 2m relative humidity.

Then $J_{synop} = J_{press} + J_{wind} + J_t + J_{rh}$

3.1 Computation of J_{press} and its gradient

Assuming p is the independent vertical coordinate, the observed quantity is actually one geopotential Z_o (either the station geopotential height, or 0 meter...) at level P_o (either the station pressure or the pressure reduced to mean sea level).

J_{press} = $((Z_o - H_v X^{mvo})/S_o)**2$ with S_o being the standard deviation of the observation error.

In this case, H_v is the post-processing operator for geopotential height which we apply to the model variables at the SYNOP point.

When computing $H_v X^{mvo}$ = geopotential value at observed pressure P_o, the components of X^{mvo} which are used are the surface pressure P_s and all the thermodynamic variables of the model. So, the gradient of J_{press} has a non-zero component with respect to these variables. This means that a pressure observation from SYNOPs has a direct impact on the humidity variables of the model, and not only on Ps and temperature. It means also that in such a variational scheme, the distinction between mass/wind analysis and humidity disappears: each observation is used to update all the model variables to which they are linked through the "post-processing " operator H_v.

3.2 Computation of J_{wind} and its gradient

The post-processing operator H_v is the one computing the 10m wind from the model variables through the flux evaluation. The inputs to this operator are the model variables at the first level and at the surface. The gradient of J_{wind} will have a non-zero component with respect to these input parameters, and it can be computed through the adjoint of the routine post-processing the 10m wind.

J_{wind} can be evaluated in two ways:

- from the wind components:

J_{wind} = $((U_o-U_{model})/S_{ou})**2$ + $((V_o-V_{model})/S_{ov})**2$ (S_{ou} and S_{ov} are the standard deviations of the observation errors for wind components)

- from the direction and speed:

J_{wind} = $((D_o-D_{model})/S_{od})**2$ + $((F_o-F_{model})/S_{of})**2$. This second formulation is in principle equivalent, but might be more convenient for some special observations, when we trust more the direction than the speed or vice-versa, when we want to use one of them only, or when we want to have a direct control of either the direction or the speed observation error.

3.3 Computation of J_t and its gradient

The post-processing operator is the one computing the 2m temperature from the model variables. We need to compute the temperature at the real observation point, so the operator H_v should include a correction (0.0065K/m) taking into account the difference between the station height and the model orography. Also stations should not be used when this difference is too large.

$$J_t = ((T_o - T_{model})/S_o)**2$$

T_{model} is the 2m post-processed temperature H_vX , and S_o is the 2m temperature observation error which includes also the error of the operator H_v (representativeness error). Two remarks must be made on the representativeness error:

- On many occasions this representativeness error dominates the instrumental error for the 2m observations;

- On many occasions it is also reasonable to assume that the representativeness errors between two different SYNOPs are correlated.

Example: land SYNOPs over a cold area with a night inversion, where the operator H_v will be affected by the same type of large error for all the different observation points. If we want also to take this correlation into account, then J_t cannot be computed independently for each SYNOP:

$J_t = (T_o-T_{model})^t\ O^{-1}\ (T_o-T_{model})$, where T_o and T_{model} are now vectors corresponding to an ensemble of SYNOPs, and O is a non diagonal covariance matrix of observation errors.

The gradient of J_t has a non zero component with respect to all the input parameters of H_v, that is the variables at the first model level and the surface variables T_s and W_s (soil moisture). Again the adjoint of the 2m temperature postprocessing operator can be used to determine the gradient.

3.4 Computation of J_{rh}

For the humidity parameter the actual measurement is the wet bulb temperature, and the quantity transmitted in SYNOP code is the dew point temperature. However it is more convenient to work out the cost function J_{rh} from the observed relative humidity (recomputed from the temperature and dew point), and from the relative humidity post-processed from the model.

Then J_{rh} = $((RH_o\text{-}RH_{model})/S_{rh})**2$, with S_{rh} the standard deviation of the RH observation error. The postprocessing operator to evaluate RH_{model} uses the temperature and specific humidity at the lowest model level and the surface pressure. It has been studied in more detail by Vasiljevic (1988).

Remarks:

- The relative humidity data will have a direct impact on the same variables as for J_t: let us note again that the mass/wind and the humidity analyses are completely simultaneous.

- As for 2m temperatures, we might introduce a horizontal correlation between humidity observation errors belonging to different SYNOPs, over one given area, in order to take into account the representativeness error of the operator H_v.

- Another possibility to investigate is the use of a correlation between the temperature observation error and the humidity observation error, as we expect the accuracy of the thermometer to have a direct impact on the accuracy of the SYNOP relative humidity.

3.5 Strategy for carrying out analysis experiments

a) STEP 1: the experiment which is close to the present operational analysis should not include J_t; it should include J_{press}, J_{rh} and J_{wind} for SHIPs and for a selection of tropical land SYNOPs. The correlation between observation errors from different data should be assumed equal to zero.

b) STEP 2: the following versions have to be tried:

- Use the 2m temperatures (include J_t in the cost-function);
- Use all the 10m winds;
- Use the direction and the speed for winds rather than the wind components.
- Introduce a positive horizontal correlation between observation errors for 2m temperatures and relative humidities.
- Introduce a cross-correlation between temperature and relative humidity observation errors (i.e. do not separate J_t and J_{rh}).
- Possibly use snow and precipitation observations to update the snow and soil moisture variables of the model.

4. COMPUTATION OF J_o FOR BUOYS

Similar to SYNOPs. Most of the time the buoys report only the pressure, so the term J_{press} is the only relevant one.

5. COMPUTATION OF J_o FOR AIREPs

The AIREPs are reporting wind and temperature data, of which only the wind is used in the present OI analysis. Then, in STEP1 of the variational analysis, only the wind data will be used. But in STEP2, a version of the variational analysis which includes the temperatures has to be tried. In addition to the wind cost function J_{wind}, it includes a temperature cost function J_t, J_t and J_{wind} being completely independent (temperature and wind observation errors are uncorrelated).

5.1 STEP1: computation of the cost function J_{wind} and its gradient

The AIREP data are normally reported at one given pressure level. To evaluate J_{wind} and its gradient, the postprocessing operator H_v is straightforward: vertical interpolation of the wind from the model level to the observed pressure level.

5.2 STEP2: computation of the cost function J_t and its gradient

The postprocessing operator is also straightforward: standard postprocessing operator for the upper-air temperature; vertical interpolation from the model levels.

6. COMPUTATION OF J_o for SATOBs

SATOBs are mainly cloud motion winds derived from the imagery of the geostationary satellites. The observed quantities are radiance patterns at different times. As it is very difficult to compute such quantities from the model variable X, it is very difficult to assimilate these quantities directly although this possibility has been studied (Eyre, personal communication).

The wind observations are chosen as input to the variational analysis like in the present OI. The postprocessing operator H_v is the same as for AIREPs. In STEP1, the different SATOB observation errors are assumed uncorrelated in the horizontal, like in the present operations. In STEP2 such a horizontal correlation has to be tried.

7. COMPUTATION OF J_o FOR TEMPs

For each individual radiosonde, the contribution to the cost-function and its gradient can be computed separately for the following parameters, as the observation errors associated with these parameters can be assumed uncorrelated:

- wind data;

- geopotential or temperature or both;

- relative humidity.

$$J_{rs} = J_{wind} + J_{zt} + J_{rh}$$

The postprocessing operators which are involved for the different parameters are the standard ones. The observation error correlation matrices which are involved in the computation of J_{wind}, J_{zt} and J_{rh} are small (size = number of observed levels).

7.1 STEP1 does not use any temperature data, but geopotential data at standard levels only in J_{zt}. J_{wind} uses also only standard levels.

7.2 In STEP2, the following versions have to be tried:

- Use of non-standard wind level data in J_{wind}.

- Use of non-standard height level data in J_{zt}: no temperature data used directly; the temperatures at characteristic levels are used to recompute extra geopotential height values.

- Insert in the computation of J_{zt} both geopotential data (at standard levels) and temperature data at all the levels when they are available.

- For the radiosondes where all parts are available, use the temperature data rather than the geopotential height data. REMARK: the need to insert an observation error correlation between different radiosondes is not foreseen.

8. COMPUTATION OF J_o FOR PILOTs

The computation is the same as for the cost function J_{wind} in a radiosonde. In STEP2, the PILOTs reporting winds at a given height level z (instead of a pressure level p) have to be treated through a specific postprocessing operator to be developed.

9. COMPUTATION OF J_o FOR SATELLITE SOUNDINGS

9.1 Interface between satellite data and variational analysis

Currently the information of the polar orbiting satellites is used as SATEM retrieved soundings in the OI analysis. However the SATEMs are not the genuine observed quantities, but an "interface" produced by a specific "retrieval" technique. An important part of the SATEM information is coming from the retrieval technique (inversion algorithm, and possibly initial profile to start the inversion with). The retrieval technique is also probably responsible for a large part of the correlation between observation errors of different SATEM observations.

In the variational context, the natural way is to avoid interfaces such as SATEMs, and to use data which are as close as possible to the observed quantity. The only restriction is that the observed data d has to be related to the model variables X through a mathematical operator H which is well defined (HX = d), differentiable (because of the gradient computations), and accurate enough (otherwise we will insert only noise in the model instead of real information). The error of the operator H is the representativeness error and should be included in the matrix O.

For satellite data the real observed quantities are raw radiances, so the target should be to use raw radiances in the variational analysis. However raw radiances are strongly affected by clouds, so the corresponding operator H would involve the model clouds to a large extent, and the quality of the present model clouds is not good enough to rely on them. Then one safe solution consists in trying to USE CLEAR RADIANCES in the variational analysis, after an appropriate cloud clearing algorithm which does not involve the model variables X at all (e.g.: the N* cloud clearing technique or the 3I algorithm does not use any forecast model - then the resulting clear radiances can be used in the analysis without any "incest problem").

Eyre (1987) chose a different approach in which the cloud-clearing algorithm is integrated with the retrieval procedure in a single variational problem. Then his cloud-clearing algorithm is helped by the temperature and humidity profile of the model. This approach is in principle better (especially if it is generalised to the 3D or 4D context) as better clouds should mean better cleared radiances and better retrieval. If the clouds are "over-constrained" in such a minimization scheme (kept close to a preliminary evaluation), it is then almost equivalent to the one using clear radiances.

The radiances are the best example of data which are linked to the model variables through an operator H_V which is not linear and rather complicated (radiative transfer equation), and which can still be used in the variational analysis. For STEP1 of the project the use of clear radiances will be considered, although it is against the general strategy defined in 1.5.

The design of the code using SATEMs is still useful for further comparisons; it is described in the annex. The main technical points treated in this annex are related to the computation of J_O when the matrix O is large and non-diagonal. Other points are technical details such as possibilities of filtering an observation bias. These points may be sometimes also relevant for the use of clear radiances, and even other observation types.

The vertical part of the variational analysis using radiances is referred to as the "1D VAR": it has to be seen as one particular module of the 3D VAR, but it could be used also to perform a 1D vertical analysis... which is a production of some kind of SATEMs. Although the initial goal is to concentrate first on the use of cloud-cleared radiances, it is better to keep the possibility of running also the more general approach when developing the variational analysis: this leaves the possibility to use also the raw radiances and to integrate the cloud-clearing algorithm in the variational system.

9.2 Computation of J_o and its gradient for clear radiances

The operator H_v which is needed to compute clear radiance values from the model variables is a direct radiative transfer model tuned to compute the radiances of the different channels available on the satellite instruments (TOVS for the moment).

For the computation of the gradient of the radiance cost function J_{rad}, we apply the adjoint T_r^* of the radiative transfer equation to derive the gradient with respect to the model quantities from the gradient with respect to the radiances (some code for the adjoint of T_r is available in France (Moll(1988)). The computations are also made in Eyre (1987)).

If we can assume that the radiances are uncorrelated in the horizontal each satellite point contributes to the gradient computations only at the 4 grid points surrounding the observation. Then the computations can be organized observation by observation, like for say the radiosonde observations. However, this assumption might not be the more realistic one as a horizontal correlation between clear radiance observation errors might come from the cloud-clearing algorithm; we need to keep the possibility of treating them as SATEMs by handling big matrices O to invert. Because the radiance computation is often affected by biases which are air-mass dependent, we need also to keep the bias-filtering possibility for the radiances as it is described for SATEMs (see annex).

As the radiative transfer equation T_r is strongly non-linear, the use of the variational technique should be a significant improvement compared to the present OI technique: the main reason is that the gradient computations will be made about the genuine model profile; only profile-dependent statistics (with an infinity of profile types!) could achieve a similar description of the link between radiances and model variables. Another implication of the non-linearity of T_r is that the cost-function J_{rad} has to be studied carefully and separately, in order to understand its behaviour with respect to the model variables. Of particular interest is the possible existence of multiple minima (see Moll (1988)).

9.3 Strategy for developments and experimentation

In STEP 1 of the experimentation program, we will use the clear radiances. In order to make sure that this research will be successful, some work has to be done on the satellite data side in parallel. This work includes:

- The best strategy for cloud-clearing algorithms (as the success of the scheme using clear radiances is highly dependent on the quality of ~~the~~ these cloud-cleared radiances);

- The determination of one suitable code for the forwards radiative transfer computations T_r, together with the adjoint code. We must try to reach a standardization of the postprocessing operator for T_r and its adjoint, like for the other postprocessing operators.

In STEP 2 of the experimentation program, the use of clear radiances will be compared to the use of SATEMs. Experiments have to be made also on the horizontal correlation of observation error for radiances, and on the techniques to remove the radiance bias (see annex). Finally some experiments have to be made with raw radiances, using Eyre's approach incorporating the cloud-clearing algorithm.

10. FUTURE OBSERVING SYSTEMS IN THE VARIATIONAL ANALYSIS

The following observations are not available yet. They are not considered for STEP1 of the development of the variational analysis. Some general ideas are given about the way they should enter the variational analysis when they become available.

10.1 Scatterometer data

The scatterometer instruments such as the one which will be available on the ERS1 satellite are observing a backscatter signal (σ_0) from the surface of the ocean. These data are sometimes used to produce some surface winds through empirical retrieval techniques.

Some preliminary studies indicate that the "σ_0" can be used in a variational analysis in the same way as the TOVS clear radiances. The operator H to examine is not the radiative transfer equation, but the operator going from the model variables to the stress, then from the stress to the "σ_0".

10.2 Future radiometers on future satellites (AMSU,AIRS)

For each new instrument a new radiative transfer model has to be used as operator H. For an instrument with a huge number of channels such as the planned AIRS, a specific study of the cost function and of the operator H will be required.

10.3 Profilers

The profilers observe (from the ground) the temperature or the wind as a function of the height: T(z) and V(z). In order to use these data in a fully consistent way, the vertical coordinate should be z rather than the pressure p (as used currently) when performing the cost function computation and its gradient. Then special postprocessing operators are needed which compute the model values of T and V at a given level z.

Because of their high observing frequency, the profiler information should become especially interesting in the context of the 4D VAR rather than the 3D VAR.

10.4 Lagrangian tracers

We may have to envisage some observations which look like the trajectory of a tracer over a long time period. To extract some useful information for the variational assimilation from this tracer, we may have to compare the observed trajectory with the model variables at different consecutive time steps. This would require having the model X not only at one time step but over a time period. This facility is not provided by the initial design.

REFERENCES

Courtier, P. and O. Talagrand, 1987: Variational assimilation of meteorologocal observations with the adjoint vorticity equation - Part 2. Numerical results. Q.J.Roy.Meteor.Soc., 113, 1329-1368.

Courtier, P. and J. Pailleux, 1987: Toward a 4D variational assimilation. RD memo R2327/1917 of 18/2/87.

Eyre, J. 1987: Inversion of cloudy satellite sounding radiances by non-linear optimal estimation: theory and simulation for TOVS. Q.J.Roy.Meteor.Soc., 113.

Gandin, L. 1988: Comprehensive quality control. ECMWF seminar proceedings on data assimilation and the use of satellite data. September 1988.

Lorenc, A. 1988: Optimal non-linear objective analysis. Q.J.Roy.Meteor.Soc., 112, 1177-1194.

Moll, P. 1988: Inversion variationnelle de radiances observees. Memoire de DEA - Universite Paris 6 (juin 1988).

Pailleux, J. 1987: Proposals for the development of a variational analysis. RD memo R2327/1965 of 23/4/87.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation - Part 1. Theory. Q.J.Roy.Meteor.Soc., 113, 1313-1328.

Vasiljevic, D. et al 1988: Use of surface observations in a variational analysis. RD memo R2304/2302 of 3/10/88.

See also in the proceedings of the ECMWF seminar on data assimilation and the use of satellite data, the part devoted to variational analysis: papers by Talagrand, Courtier, Lorenc, Pailleux, Betout (September 1988).

ANNNEX: SPECIFIC COMPUTATIONS RELATED TO SATEMs

In the computation of J_o and its gradient a specific problem is raised for the SATEMs because their number can be huge, and their observation errors are normally correlated in the horizontal, leading to huge matrices O. The present annex deals with this specific problem.

Although it is not envisaged to insert the SATEMs as such in STEP1, the code will have to be developed at some stage, because:

- In STEP2 we want to compare the use of radiances and the use of SATEMs;

- The treatment of big matrices O has to be considered also for radiances and probably other observing systems; the technique described thereafter for splitting the vertical and horizontal computations is general;

- The bias removal technique, described thereafter on the SATEM example, can be also generalized.

A. Postprocessing operator for SATEM thicknesses

The operator H_V needed to compute the thicknesses from the model is the post-processing operator for the geopotential height applied to the bottom and top level of the layer.

B. How to split the J_{satem} computation

Because of the horizontal correlation between SATEM observation errors, O contains many non-diagonal terms, and we cannot separate the contribution of each individual SATEM as we generally do for all the other obs types.

However there are physical reasons to split J_{satem} in a sum of different contributions J(I1,I2,I3,I4),with:

I1 = 1,N1 (number of satellites)

I2 = 1,N2 (number of retrieval types - presently N2=3)

I3 = 1,N3 (number of latitude bands or air masses - N3=6 with the NESDIS statistical retrieval working on 6 different latitude bands. N3 could be 1 with a consistent retrieval technique).

I4 = 1,N4 (number of surface types - presently N4 could be 3: water, ice and land.)

To split the computation means that the correlation is zero whenever the two SATEMs do not belong to the same category (same satellite, same retrieval type,...). This assumption is questionable in some cases: for example in the stratosphere, we should expect a positive horizontal correlation between two SATEMs one being over sea and the other over land, as in a good retrieval technique the stratosphere must not be contaminated by the surface properties. However this assumption is not a big problem, taking into account the fact that the assumption of a correlation depending only on the distance looks much more questionable.

The separation of the computations follows the general logic of having a maximum number of diagnostic figures (see 1.1). For example we might find that the cost function of the MSU soundings on a given area is too high compared to the corresponding cost function of the clear soundings, due presumably to a bias in the retrieval, and we might want to do something to cure the problem.

One way to cure the bias problem is illustrated by the following example: suppose that for one given satellite, we think we have a large bias on the MSU SATEMs for one given area of the globe. Then provided we have enough other observations on this area (clear SATEMs or radiosondes), we let the mean value of X being driven by these observations, and we filter the bias of the MSU soundings by using for the cost function J_{msu}:

$$(E-M)^t \, O^{-1} \, (E-M) \text{ instead of } E^t \, O^{-1} \, E$$

M is the mean value of the normalised departures for the MSU computed on the sample corresponding to the given satellite and the given area.

C. <u>Computation of one SATEM cost function J(I1,I2,I3,I4):</u>

With the previous splitting and the present volume of data, each term of the cost function should not involve more than N_{max} = say 700 SATEMs. If however this number is exceeded we will put an extra arbitrary separation according to the observation time; basically the data of one orbit would be split in segments in a way similar to the one used in a satellite retrieval procedure.

Then we are faced with the computation of a cost function equal to $E^t O^{-1} E$, where the matrix O corresponds to up to 700 SATEMs with up to 7 layers in the vertical. It is not easy to invert O which would be as large as 4900 X 4900! Fortunately a specific technique can be worked out by taking into account the usual decoupling assumption of the correlation in one vertical correlation and one horizontal correlation.

Let us call V the vertical correlation for the obs error of the L layers (L = 7 normally), and M the horizontal correlation matrix for the N SATEMs (N up to 700). We can see that, due to the specific shape of the matrix O, the quadratic form

$J = E^t O^{-1} E$ can be computed in the following way:

(i) For each SATEM k = 1,N compute the L-component vector $P_k = V^{-1} E_k$, E_k being the vector of the L thickness departures (normalised) for SATEM number k.

(ii) For each layer l = 1,L compute the N-component vector $Ql = M^{-1} El$, El being the vector of the N SATEM departures (normalised) for layer number l.

Remember that in a typical case the dimension of E_k and P_k is L=7, and the dimension of El and Ql is N=700.

(iii) Let us call P the vector obtained by putting together the different vectors P_k:

P =[P1]
[P2]
⋮
[PN]

Let us call Q the vector obtained by putting together the different vectors Ql:

+ =[Q1]
[Q2]
⋮
[QL]

and then reordering the elements in the way similar to P: L layers of first SATEM, L layers of second SATEM,....

(iv) Then the cost function we want to compute is $J = P^t Q$, that is the dot product of the two vectors P and Q.

This technique is completely equivalent to a projection of E_k on the normal modes of the matrix V, before a computation of the cost function for each individual mode using the matrix M (it might actually be useful to separate the contributions to the cost-function of each individual vertical mode and to store them as interesting diagnostic quantities).

Remarks:

The previous computations can be set up easily if we make the following assumptions:

- For one SATEM category the number of levels we use in the analysis is the same;

- Whenever one layer is rejected, we reject the total sounding, in order to avoid having incomplete soundings. This assumption is not done at the moment but is reasonable.

D. Computation of the gradient of one SATEM cost-function

The computation of the gradient of one cost function $J = E^t O^{-1} E$ can be done through the following steps:

- $\mathrm{Grad}_{(E)} J = 2 O^{-1} E$: the same kind of computation as on J is required as O is not diagonal (use of the eigenvectors of V). We need a table as large as the total number of SATEM data to store this gradient. However all the remaining steps of the gradient computation can be done individually for each SATEM profile without any big table to store in central memory.

- Computation of the gradient with respect to the vector $H_V X^{mvo}$ which in this case contains the values of the SATEM thicknesses post-processed from the model.

- Computation of the gradient with respect to the vector X^{mvo}.

- Computation of the gradient with respect to X by multiplication with the coefficients of the bilinear interpolation in the horizontal.

The last three steps are the standard ones described in 1.4.

E. Postprocessing operators for SATEM PWC

For SATEM PWC the postprocessing operator H_V is the integral of the specific humidity q on the observed layer, using p as independent variable in the vertical.

In the present humidity analysis the PWC observation errors are assumed uncorrelated both in the vertical and in the horizontal. To reproduce this assumption in the variational context means that the SATEM observations can be treated one by one (like the non-SATEM observations) for the computation of the cost function and its gradient. The reason is that the matrix O is diagonal.

However the purpose of the variational analysis is more general than a pure reproduction of the present system, and it is better to keep open the possibility of a correlation between the PWC obs errors. This means that the organisation of the cost function for PWC has to be organised in a way very similar to what we do for thicknesses, with all the SATEM information in memory.

Another option which could be tried easily in the variational context consists in assuming a non-zero correlation between the PWC and the thickness obs error. Then the contributions of PWC and thicknesses to the cost function cannot be separated; the computation is similar to the one described before except that the size of the matrix V is L = 10 instead of 7: 7 thicknesses + 3 PWC.