WORKSHOP ON NON-LINEAR ASPECTS OF DATA ASSIMILATION: Reports of Working Groups

1. WORKING GROUP 1: ASSIMILATION METHODS

Although the workshop was devoted to nonlinear aspects of data assimilation, the working group concentrated on extensions of the quasi-linear framework where the tangent-linear approximation remains valid. The analysis has to be correct in the linear case before much work is devoted to non-linearities.

1.1 Carrying information from one assimilation window to the next, taking non-linearities into account

What is being done today at ECMWF? The atmospheric state evolves according to the prediction model. The second order statistics are stationary for the correlations, and the variances evolve according to a simple law. The humidity variances remain based on a simple empirical formulation. In some other centres, e.g. UKMO, some empirical dependence on the basic flow has been introduced.

At a fundamental level, there are two ways of representing the probability density function (pdf) of the estimate: parametric or sampling. Both frameworks are useful and complementary. The sampling approach is in general independent of any linearity hypothesis.

This is a field of rapid development, the long-term evolution of which may take various routes. ECMWF should cooperate with member states, universities and research laboratories to follow and support research activities in this wide and open field. The situation could be summarised as follows. Many methods for describing the temporal evolution of the uncertainty of the estimation of the state of the flow have been proposed and are being studied. It is still too soon for ECMWF to identify in which procedures to invest heavy resources. ECMWF should nevertheless continue its promising research effort in the subject.

1.1.1 Variances

What is "easily" feasible? With the tools already developed at ECMWF, a simplified deterministic prediction has been implemented to produce an index of the spatial variability of the variances. A similar achievement may be obtained using the Ensemble Prediction System. Simple dynamical models (e.g. relying on potential vorticity dynamics) may already provide a significant amount of useful information.

1.1.2 Correlations

The static correlations implemented in most operational centres for describing the short-range forecasts errors have led to robust assimilation systems which have been shown to work reliably and satisfactorily in practice. As a consequence, any flow dependences are likely to be implemented as an "add on" on top of the static description.

Experiments performed so far in simplified but meteorologically relevant contexts have shown that the spatial correlation scale of the forecast errors increases with time whereas the analysis, particularly in data dense areas, tends to reduce this length scale.

Investigation and documentation of the time evolution of the correlations in a data assimilation cycle should be pursued by ECMWF. The simplified Kalman filter under development at ECMWF should improve the specification of these correlations in baroclinic areas.

Orography, and other stationary forcings, may induce non-homogeneous and non-isotropic correlations which in general have both static and flow-dependent components. This too should be investigated by ECMWF.

1.2 Identification and description of non-Gaussian behaviour of relevant pdf's

Today, there is no description of bimodality in the analyses produced by operational centres, whereas in the future it is likely that the result of the assimilation will be a richer description of the pdf than just a mode or an expectation and the associated variance. More generally, current assimilation systems assume that background errors are approximately normally distributed. The degree to which this assumption is valid should be tested.

Quality control, and specific observations like the wind scatterometer, may induce bi- or multi-modality. Even if the analysis errors are elliptical in phase space, the dynamics may generate bimodality.

Is it possible to diagnose forks (splitting of a unimodal pdf into several modes)? Appropriate OSSE's (generating a sample according to the specified observations errors) would help in assessing this issue, and would start to produce quantified results. The results may be classified according to weather regimes and/or meteorological flow patterns.

Description of forecasts errors in terms of synoptic features (e.g. misplaced frontal structure, or overdeveloped low) is likely to generate progress in the conceptual formulation of assimilation methods.

1.3 Adaptive estimation of the relevant statistics

It is not possible to estimate everything. Only a few parameters can be evaluated robustly from the sequence of innovation and residual vectors. As a consequence, it is necessary to have models which reduce the problem to the adaptive estimation of a few parameters of both the dynamics and the statistics.

The HIRLAM group has been investigating the issue, and the approach has been used for a single case. Results were presented at the seminar by GRGS, Toulouse in an oceanic model under the assumption of a stationary Kalman gain, with further developments being still to come. Other groups are, for example, comparing several adaptive schemes in data assimilation for the tropical Pacific ocean.

Certain properties of a posteriori statistics can be used to diagnose flaws in the assumptions of the assimilation. In the variational context, several diagnostics mentioned during the Seminar and the workshop, even some very elementary ones like the χ^2 distribution of the objective function at the analysis point, can be implemented and should lead to some form of quantitative diagnosis of weaknesses in the specification of key parameters. As a further example, the adjoint trajectory at the minimum is the Lagrange multiplier associated with the strong constraint of the dynamics. As such, it contains information on the model error.

1.4 Chemical constituents and ozone assimilation

The use of ozone and other variables whose physics is dominated by advection will contribute to the estimation of the wind field. Error statistics will have to be specified according to their probability law of error, which may be more log-normal than normal. The assessment of the validity of the tangent-linear hypothesis will have to be performed with respect to the data availability.

In the longer term there is also potential for extracting temperature information from chemically active species, as well as wind information at sunset and sunrise. This information would complement what is extracted from passively advected species. It should nevertheless be noted that the chemistry models have generally not been validated at the global scale.

Real time availability of ozone data, and more generally of data on other trace species, will be needed for ECMWF to become fully involved in this field. ECMWF should liaise with the appropriate space agencies to ensure this.

1.5 Assimilation of imagery, in particular from geostationary satellites

The assimilation of radiances from geostationary satellites under clear conditions should help determine the dynamical fields through their temporal dimension.

Pattern recognition is not, so far, a standard tool in data assimilation. ECMWF should appraise itself of what is being done on this topic in other fields, including physical oceanography.

1.6 Assimilation of land-surface variables

A significant amount of information from the SYNOP data is not being used currently. Difficulties related to the representativeness of the measurements will have to be overcome; the introduction of subgrid scale open water is a step forward in this direction. This is a field in which some member states are more advanced, and significant progress in it should be made by ECMWF.

It is not recommended at this stage to investigate a fully coupled soil/atmosphere assimilation. Major improvements are likely to arise from a stand-alone soil analysis consistent with the boundary layer.

1.7 Coupled ocean/atmosphere data assimilation

The very different timescales involved in the deep ocean compared to those of the atmosphere will introduce new challenges for data assimilation. Up to now the two analysis problems have been mostly decoupled.=

Fundamental research on a "slow manifold" concept for the coupled ocean-atmosphere is probably required, but effort for developing a constructive algorithm has produced limited results so far.

1.8 Perturbations for ensemble prediction

Assimilation and ensemble prediction are two facets of the same problem, namely to estimate the pdf of the state of the flow, given the observations and a dynamics. In the future these two activities will become more and more connected.

The choice of the initial perturbations of the EPS is intimately related to the analysis errors, but depends on the goals the EPS is supposed to achieve. In particular, it depends on whether the goal is to produce some possible extremes (or to bound the pdf) or whether it is to produce a sample of the pdf, including quantitative estimates that extreme events will or will not happen.

ECMWF is building its sample according to what will happen while NCEP and AES build their samples according to what has happened. An objective assessment of the different approaches with respect to the goals of the EPS is recommended.

In principle, the a-posteriori estimation of the analysis error should be used for determining the initial perturbations of the EPS accounting for their likelihood. It is probable that enhancements of the background formulation will be necessary in baroclinically active areas for achieving this goal. It should be noted that this is consistent with the current ECMWF research effort with the simplified Kalman filter.

1.9 Re-analysis and observing system (simulation) experiments

Diagnosis similar to that discussed in section 1.3 should be continued to document the reanalysis performance in extracting information from the observations and producing a posteriori error bars.

OSE's should be performed to evaluate the capability of 3D-Var to extract information from the various components of the current observing system.

OSSE's contribute to the definition of future observing systems and ECMWF has a significant role to play in this area.

1.10 Sharing of resources between the assimilation and other components of the forecasting system

The sharing of resources should rely on some objective assessment. Computer resources are available to perform numerical experimentation along these lines. However, an objective measure of what is brought by a given system to the member states remains to be defined.

2. WORKING GROUP 2: REPRESENTATION OF PHYSICAL PROCESSES IN VARIATIONAL DATA ASSIMILATION

2.1 General considerations

2.1.1 Importance of the physics

The weaknesses of adiabatic tangent linear (TL) and adjoint (AD) models have been clearly identified in the context of both variational data assimilation and singular vector computations. At ECMWF, the assimilation of SSM/I total column water vapour data using a 4D-Var approach with an adiabatic model led to unrealistic supersaturation in the analyzed moisture field. The computation of singular vectors with an adiabatic model produces rapidly-growing non-meteorological structures due to the lack of surface friction.

Non variational methods have also pointed to the importance of an explicit consideration of physical processes in data assimilation (e.g. physical initialization from Krishnamurti et al.).

The importance of including physical processes (particularly the moist convection) has also been indicated by F. Rabier using the ECMWF 4D-Var incremental assimilation system. She found that the scores of the forecasts in the tropics from 4D-Var analyses produced using very limited physics in the inner model integrations are poorer than with the 3D-Var system.

The 4D-Var approach can use the model equations as a strong or weak constraint. Thus increasing the realism of the model should lead to a better analysis. Another justification for including physics in variational data assimilation is the possibility of extracting information from new types of observations related to the parametrization of moist processes (precipitation, radar reflectivity, precipitable water). Including the physics of surface processes should also be beneficial for the assimilation of surface and satellite radiance data that are sensitive to, for example, skin temperature.

The current analysis of water vapour as well as the temperature and wind profiles in the boundary layer could benefit particularly from the inclusion of physics. The importance of planetary boundary layer effects on the shape of structure functions evolved in the 4D-Var incremental system close to the surface has also been demonstrated.

However, the strong non-linearities and threshold processes associated with physical parametrizations can make the validity of the tangent linear approximation questionable leading to problems associated with non-quadratic cost functions.

2.1.2 Status of current studies

Various meteorological centres have implemented linear physics in different ways. The 4D-Var incremental approach developed at ECMWF allows physical processes to be included progressively in the linearized versions of the model. At NCEP, an approximate gradient is computed using an adjoint model including only part of the physical processes.

The adjoint of important physical processes, given the time scale of interest (less than one day), has been developed (turbulent processes, moist convection, large scale condensation) both in mesoscale models (NCAR, NCEP) and in global models (FSU).

The explicit computation of the Jacobian matrices is an alternative approach avoiding the explicit coding of adjoints (R. Errico). This approach, although not suitable in an operational context, can provide useful information on the behaviour of physical parametrizations in a development phase.

The utility of the tangent linear approximation has also been examined by comparing perturbations evolved with both the tangent linear model and the non-linear model. When the dynamics is dominant, the tangent linear approximation with physics is reasonable, but there are examples showing that when convection is the dominant mechanism the tangent linear approximation fails.

Results shown during the Workshop demonstrate that the inclusion of the physics in a 4D-Var assimilation as well as the assimilation of variables such as precipitation or precipitable water can have a positive impact on the quality of the analysis for moisture and precipitation. This has been demonstrated for assimilation periods ranging from 3h to 12h with limited area models (X. Zou and D. Zupanski) and global models (T.Tsuyuki).

For these case studies, non-linearities and/or thresholds of the physics did not seem to affect detrimentally the convergence of the descent algorithm, although regularizations have been necessary in some situations.

The inclusion of the physics is nevertheless computationally expensive.

2.2 Future developments

The inclusion of physical parametrizations in 4D-Var assimilation and other applications using adjoint models (singular vectors, Kalman filter) is expected to have a beneficial impact on the quality of the products generated by operational meteorological centres (analyses, forecasts). Improvements should be more important in tropical regions where convection plays a crucial role. The new generation of computers will allow the extra computational expense (memory storage, number of iterations). Some regularizations may be necessary but being model-dependent, no general rules can be given. In particular, simplifications or regularizations of the physics may be less damaging in 4D-Var incremental assimilation than in full 4D-Var assimilation.

New types of observation will be assimilated (radar data, raingauges, GPS data), and some of the current observations should be better used (clouds, surface observations). This will need improvements of physical parametrizations (e.g. the use of prognostic cloud schemes instead of diagnostic cloud schemes) since they will be part of the observation operator, and may increase the need for some accounting for model error in the data assimilation process. New control variables related to the hydrological cycle (cloud liquid water) and to surface processes (soil moisture, skin temperature) could be added in 4D-Var assimilation.

2.3 Remaining problems

The following issues have not yet been addressed:

- More accurate (and computationally more expensive) physical parametrizations may need to be developed for assimilation of some types of data than can be justified for medium-range forecasting purposes.
- Some of the current physical parametrizations may have to be modified to make them suitable for tangent linear applications.
- The consequences for minimization properties of non linearities including thresholds are not adequately understood.
- It is not clear how to deal with non-Gaussian probability distribution functions for observational and background errors (for example with regard to precipitation).
- The use of gravity-wave control may become a different issue in the presence of physics due to interactions between the processes, for example the effects of moisture convergence on convection.

2.4 General recommendations

The working group recommends:

- Implementation of a more realistic set of physical parametrizations in the linear versions of the ECMWF model used by the 4D-Var assimilation system.
- Improvement of the physical parametrizations with regard to their applications in data assimilation (accuracy, tangent linear approximation, representativeness of observed quantities).
- Carrying out developments in physical parametrizations in collaboration with data assimilation developers and in the light of what will come from new observing systems.
- Investigation of diverse set of diagnostics for evaluating the impact of physics on variational assimilation (Jacobian elements, measures of the utility of the tangent linear approximation, measures of gravity wave activity, various measures of both local and global skills).
- Improvement of the specification of background and model errors using some input from the physical parametrizations.
- Study of the effect of non-linear observation operators and non-Gaussian error statistics on the minimization process.
- Following of developments on automatic adjoint coding techniques and new numerical minimization methods for non-linear problems of large dimension. Evaluation of these new tools when they become available.

3. WORKING GROUP 3: OBSERVATION OPERATORS, ERROR STATISTICS ANDQUALITY CONTROL

3.1 Strategy for improved exploitation of observations

NWP centres which have recently implemented, or are implementing, variational data assimilation systems are now considering how best to exploit the possibilities provided by these new systems for improved use of observations, including improvements in: monitoring, pre-processing, observation operators, characterisation of errors, quality control, treatment of biases, analysis of humidity, ozone and surface fields, and exploitation of new observation types. However, resources available to address all the potential areas of improvement are limited, and so **it is recommended** that centres improve their collaboration in this area, including exchange of plans, documentation and software modules. As a starting point, it would be useful to establish, for each observation type, the scope of present and planned work on observation operators, error characterisation, etc. The optimal design of the observing network for NWP is another important topic but this was considered to be beyond the scope of this workshop.

3.2 Observation operators

Efficient and sufficiently accurate observation operators are needed for all the types of data used in the assimilation. The formulation of observation operators must take into account any pre-processing that has been applied to the measurements. For this reason **it is recommended** that regular liaison be maintained between the users and the data providers both on the existing pre-processing and to provide input for developing better pre-processing methods. For each element of the observing system **it is recommended** that the transmission of all the necessary information required for the observation operator should be included in the data (e.g. for radiosondes instrument type and time/location at each level should be specified). **It is recommended** that agencies responsible for satellites and other remote sensing measurements for NWP applications should support the development of observation operators, in liaison with NWP centres, to improve the utilisation of the data from existing and new instruments.

With many new forms of data (e.g. surface observations, total column ozone, geostationary satellite radiances, GPS radio-occultation data) becoming available it will be increasingly important to co-ordinate the development of observation operators between NWP centres since no single centre can afford to develop all of them. This implies the code for the operators should be modular and well documented. **It is recommended** that improved operators for radiosondes and other conventional observations be developed so as to make better use of the basic observing systems. Different observation operators may be required for different instrument types (e.g. radiosondes). In a minimization with non-linear operators (e.g. TOVS) there may be a trade-off between the density of observations which can be processed and the number of iterations of the forward model calculations. This problem may be reduced by using a tangent linear approximation with occasional relinearization around a complete nonlinear forward model solution. The frequency of the relinearization would be observation-type dependent and would have to be determined by future research. The need for more observations will be a function of the variability of the field being inferred (e.g. cloud cover will require a high observational density). **It is recommended** that this be investigated.

3.3 Observation error statistics and properties

A unified approach should be taken towards the estimation of relevant observation error characteristics for all observing systems. It is recognised that there is a significant effort needed, both for the scientific work and technical implementation of an automated system for deriving and updating observation error statistics. For this to be successful and robust, **it is recommended** that there be a continuous long term effort in this area as part of the data assimilation development work. The system needs to be regularly reviewed both for scientific impact and for technical robustness. There should be automated checks of consistency between analysis feedback statistics and the specified observation error statistics. Furthermore all specifications must be saved as this information is valuable to exchange between centres and to feed back to data producers, although it should be noted that the specified observation errors represent both instrumental errors and errors of representivity.

The forecast model background fields can be used as a reference for comparing with observations, both for determining biases of individual observing systems (or subsystems) and for comparing different systems or subsystems. An automated system should be devised for computing and updating biases. However, care has to be taken not to risk a drift in a bias correction system due to model errors. A fairly large number of conventional observations will have to be left uncorrected and serve as a reference for the system. From monitoring it is well known that there is a subset of, for example, radiosondes which have consistently small biases and generally perform well.

The bias correction of satellite and conventional data (as well as the background fields) should result in improvement in the usage of observations in NWP models. A persistent and identifiable bias in the innovation vector because of processing, calibration and forward model errors has been found in satellite data. The removal of this error (which is often as large or larger than the signal) was necessary before the data could be successfully incorporated in analysis systems. Similarly, biases (sometimes less consistent in time than is the case with the satellite data) have also been found in conventional data because of radiation effects on the instrument, improper station height assignments or for other reasons. While their magnitude is in most cases smaller than the signal, the usage of the information in the observations can be enhanced by removing such biases. The best way to remove biases (for both conventional and satellite data) which are understood (e.g. calibration problems) is to eliminate them at source. However, for some data this is not feasible in the short term and an empirical bias correction is necessary. While the identification of biases should be a priority for both the data producers and users, **it is recommended** that any required empirical bias correction be applied at the NWP centres where the model background field and other observations are available.

While biases are the first priority to address, other parameters should be estimated routinely on a station by station basis. Standard deviations of observation errors can be estimated and used individually for each instrument type and probably also for groups of stations and eventually on a station by station basis. It is also desirable to estimate the shape of the probability distributions. In this context individual rejection limits ought to be derived and specified on a station basis (e.g. HIRLAM grey list).

The other component of the observation operator error is the representativeness error due to scales unresolved in the model. **It is recommended** as a second priority to estimate and make use of representativeness error in the assimilation. This will be model-resolution as well as situation dependent.

3.4 Quality control

Data assimilation systems use two types of quality control (QC) to identify erroneous data. A gross check against the background is an efficient way of identifying corrupt observations. A simple way to enhance the background check would be to make the rejection limits dependent on the probability of gross errors. This would be particularly useful in cases when there are no independent data in the vicinity of the observation to be checked.

The second, and more selective, step in the QC chain is the variational quality control (VQC) or its equivalent in OI. The efficiency of VQC depends critically on the availability of nearby independent information. An essential ingredient of efficient QC schemes is an accurate specification of the background and observation error covariances. Failure to account correctly for the non-Gaussian nature of observation and background errors may result in over optimistic estimates of analysis error variances. The VQC provides a Bayesian framework for more sophisticated QC with non-Gaussian error distributions. However, careful specification of the probability of gross errors is necessary. Quality control of dependent data increases the complexity of VQC.

The ECMWF analysis system rejects a rather small amount of data. **It is recommended** that experiments with a decreased and a significantly increased level of data rejection be carried out to test the importance of QC. A general validation of the QC procedure **is also recommended.**

3.5 Humidity analysis

There is scope for substantial improvement in the analysis of humidity. There remains uncertainty as to the best variable for humidity analysis: specific humidity, its logarithm and relative humidity are currently used. Related to this is the problem of background error and correlation, particularly between the humidity and temperature fields, where current assumptions are rudimentary. Within the variational framework the problem of the upper and lower bounds in the humidity should be addressed through the appropriate choice of analysis variable and/or through additional terms in the cost function. Effective use of humidity observations is also hampered by the problem of model bias, which should be addressed as a model problem but could also be addressed as an analysis problem. Improved analysis of humidity is likely to benefit from increased attention to the analysis of related variables such as cloud and precipitation. There is a wealth of under-exploited information in satellite and radar imagery; work is required on appropriate observation operators, and this problem is closely linked to those of physical initialisation and the adjoints of physical processes.

3.6 Ozone analysis

There is increasing interest in the extension of data assimilation systems to include ozone as an analysed variable, stimulated not only by the need for an improved description of the ozone field itself, but also by the potential through 4D-Var for improved wind analyses in the lower stratosphere. A further benefit should be an improved use of data from the HIRS temperature channels, which have some sensitivity to the ozone distribution.

ECMWF will be active in this area through its participation in the EU-funded SODA project. The principal observational source will be the HIRS channel 9 radiances; other possible sources from current satellites are the total-ozone data from TOMS on ADEOS and retrieved profiles from the short-wave measurements from GOME on ERS-2 and SBUV on the NOAA satellites. Ozone data will become available in the future from instruments on ENVISAT, EOS, MSG and METOP. Problems of bias both in the model and in the observations are likely to need particular attention, and observations from the WMO-sponsored ozonesonde and total-ozone measurement networks will be needed as ground truth. Another important task will be to determine the correlations between the background errors in ozone and those of other analysed variables. Options for the initial approach are either to assume no correlation, or to assume that the error correlations follow the known correlations between full fields.

It is recommended that an operational system for the exchange of ozone measurements be established so that all relevant satellite and ground-based data are made available to NWP centres in real time.

3.7 Surface observations

Surface (P_s precipitation, snow depth, SST etc.) and near surface (T_{2m}, RH_{2m} and $\mathbf{V}_{10m}$) observations are the most widely available ground-based observations. However, they are, with the exception of P_s, either not used or used in a limited form (e.g. $\mathbf{V}_{10m}$ over sea) in the variational analysis. The main reasons for this are: highly complex and non-linear observation operators, model surface variables not being part of the control variable and a lack of knowledge of the observation errors.

The vertical gradient of the model variables varies strongly in the lower part of the PBL due to physical factors (turbulence, terrain characteristics etc.). The observation operator should take account of the model surface layer by including the physical parametrization of the forecast model. Not having the model surface variables as a part of the control variable is a severe limitation when trying to use the surface/near- surface observations. This, for example, in the case of using T_{2m} over land can lead to unrealistic changes in the surface-layer stratification. Including the surface variables in the control vector will also lead to a better use of the satellite radiances.

The observation-operator error includes the representativeness error which may be predominant in the case of surface/near surface observations. Because of this one can expect errors of the adjacent observations (e.g. T_2) to be correlated in certain weather conditions (e.g. winter inversions over land). It should be emphasized that the representativeness error (the local sub-grid variability) must be included in the observation operator error.

In spite of these difficulties, because of the dense distribution of surface observations and their relatively high frequency in time, further work to use them at ECMWF **is recommended. It is also recommended** that national agencies make available in real time hourly data and more frequent automatic observations to NWP centres which will exploit them in the 4D-Var systems under development.