Assimilation of Satellite data for Numerical Weather Prediction: Basic importance, concepts and issues

Adrian Simmons
European Centre for Medium-Range Weather Forecasts

1. Introduction

Satellite-borne instruments form a major component of the meteorological observing system. They provide most of the volume of observational data presented each day to operational weather forecasting systems, and their data are of increasing importance in determining the accuracy of the forecasts that are produced. The goal of making effective use of satellite data has nevertheless been a major challenge to numerical weather prediction centres over more than two decades, and has been an important stimulus for the development of better methods of data assimilation.

This account of the assimilation of satellite data in numerical weather prediction begins with a discussion of the importance of the data, based both on the evolution of skill of operational weather forecasts and on the results of controlled forecasting experiments in which data from specific components of the observing system are withheld from the assimilation system. This is followed by illustration of the data assimilation problem in general and identification of some of the specific problems that combine to make the exploitation of data from satellites especially challenging. Foremost amongst these is the fact that the satellite-borne instruments do not measure directly values of wind, temperature or humidity that may be used straightforwardly to correct corresponding values in the assimilating atmospheric model. This was one of the main motivations for the development of the variational method of data assimilation, and the one-, three- and four-dimensional forms of this method (1D-Var, 3D-Var and 4D-Var) are thus described in some detail. Examples of some specific problems in the areas of data monitoring, quality control and bias correction are presented in the final section.

Several of the topics introduced in this paper are discussed more fully in a series of lecture notes from the ECMWF Meteorological Training Course on Data Assimilation and Use of Satellite Data, which may be viewed or downloaded from ECMWF's public website (http://www.ecmwf.int).

2. Improvements in operational forecasts

Fig.1 presents running annual-mean anomaly correlations of 500hPa height for ECMWF's operational three-, five- and seven-day forecasts for the extratropical northern and southern hemispheres for the period from January 1980 to February 2001. The general upward trend of the curves indicates a progressive improvement in forecast quality over the more than twenty-one years for which records are available.

Fig. 1 shows clearly a much higher overall rate of improvement in the forecasts for the southern hemisphere than in those for the northern hemisphere. In the early 1980s, the skill levels of the three- and five-day forecasts in the southern hemisphere were only a little better than those of the five- and seven-day northern hemispheric forecasts. At the time this was not regarded as a surprising result, in view of the sparsity of conventional ground-based and aircraft observations in the southern hemisphere (Bengtsson and Simmons, 1983). Today, however, the skill level at a particular forecast range in the southern hemisphere is only a little lower than that at the same range in the northern hemisphere. Although lower intrinsic error growth rates in the southern hemisphere (Simmons et al., 1995) may

compensate for a lower accuracy of initial analyses due to poorer data coverage, improved usage of satellite data, particularly over the oceans, is likely to be an important contributor to the greater rate of advance in forecast skill in the southern hemisphere.

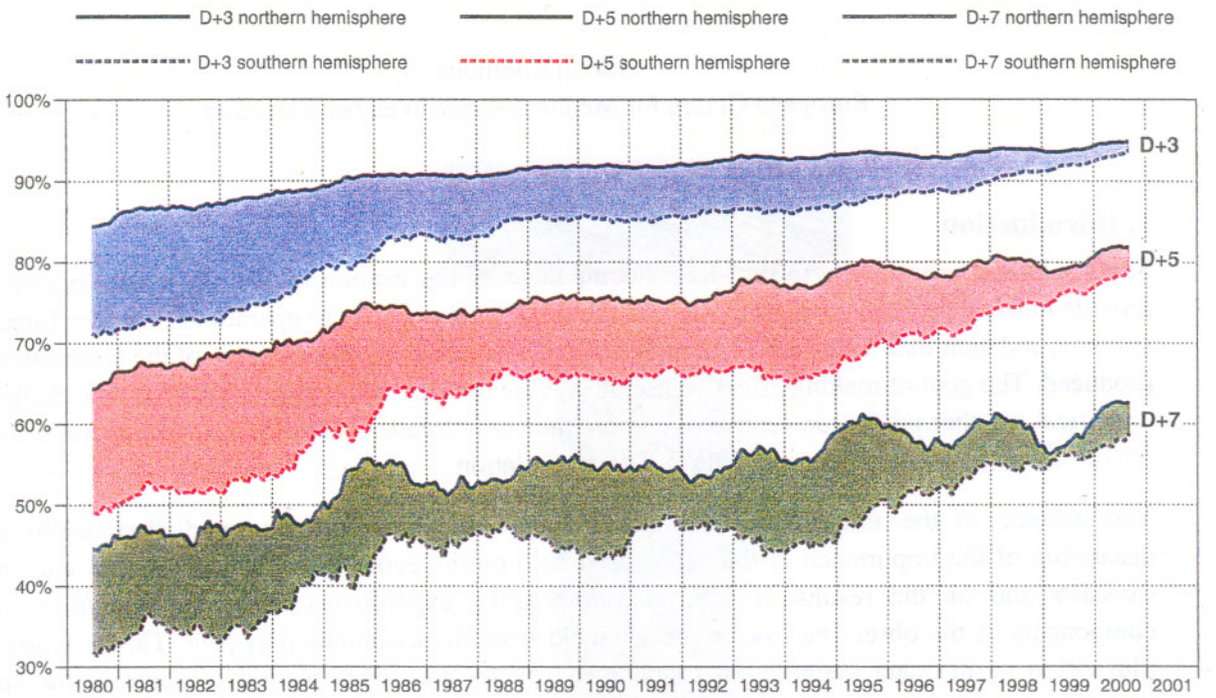

Fig. 1 Anomaly correlations of 500hPa height for 3-, 5- and 7-day forecasts for the extratropical northern and southern hemispheres, plotted in the form of 12-month running means of archived monthly-mean scores for the period from January 1980 to February 2001.

There is little doubt that improvements in the availability, accuracy and assimilation of satellite data have been major factors contributing to the overall improvement in forecast skill. New data types that have become available in recent years include information on wind and humidity over the oceans from the DMSP SSM/I instruments and information on marine winds from ESA's ERS scatterometers. Improved microwave soundings (more readily used over ocean than land) have come from the AMSU instrument first flown in 1998, and there have been evolutionary improvements in the wind products derived by EUMETSAT and other satellite agencies by tracking features in successive images from geostationary satellites. In addition, improvements have been made in the assimilation of satellite data by NWP centres, most notably the direct variational assimilation of radiances from the (A)TOVS instruments on the NOAA polar orbiters. Such developments would be expected to improve forecasts more in the southern than in the northern hemisphere, both because satellite data provide a relatively more important component of the observing system in the southern hemisphere and because of the greater extent of the oceans in that hemisphere.

The root-mean square errors of three- and five-day forecasts of mean-sea-level pressure for the two hemispheres from a number of global forecasting systems are presented in Fig. 2. Annual running means of scores exchanged under the auspices of WMO since mid 1987 are plotted. The figure shows a general trend towards lower forecast errors. A noteworthy feature of the ECMWF results for the southern hemisphere is an increased rate of improvement beginning in 1997, especially evident at day three. The starting point for this was the operational introduction of 4D-Var data assimilation (Mahfouf and Rabier, 2000, and refs.) in November 1997, which starts to affect the annual running mean in Fig. 2 six months earlier with the chosen plotting convention. Factors known from pre-operational testing to have brought subsequent improvement include assimilation (from May 1999) of raw (level-1c) microwave radiances from the (A)TOVS (AMSU-A and MSU) instruments rather than pre-processed TOVS (MSU, HIRS

and SSU) radiances (McNally et al., 1999), assimilation of surface wind-speed retrievals from SSM/I and general post-implementation refinement of the quality control and data selection of the raw-radiance assimilation scheme.

Also noteworthy in Fig. 2 are the substantial recent improvements in forecasts for the southern hemisphere produced by the Met Office and NCEP, both of whom currently use direct (three-dimensional) variational assimilation of (A)TOVS radiances (Bell et al., 2000; English et al., 2000;

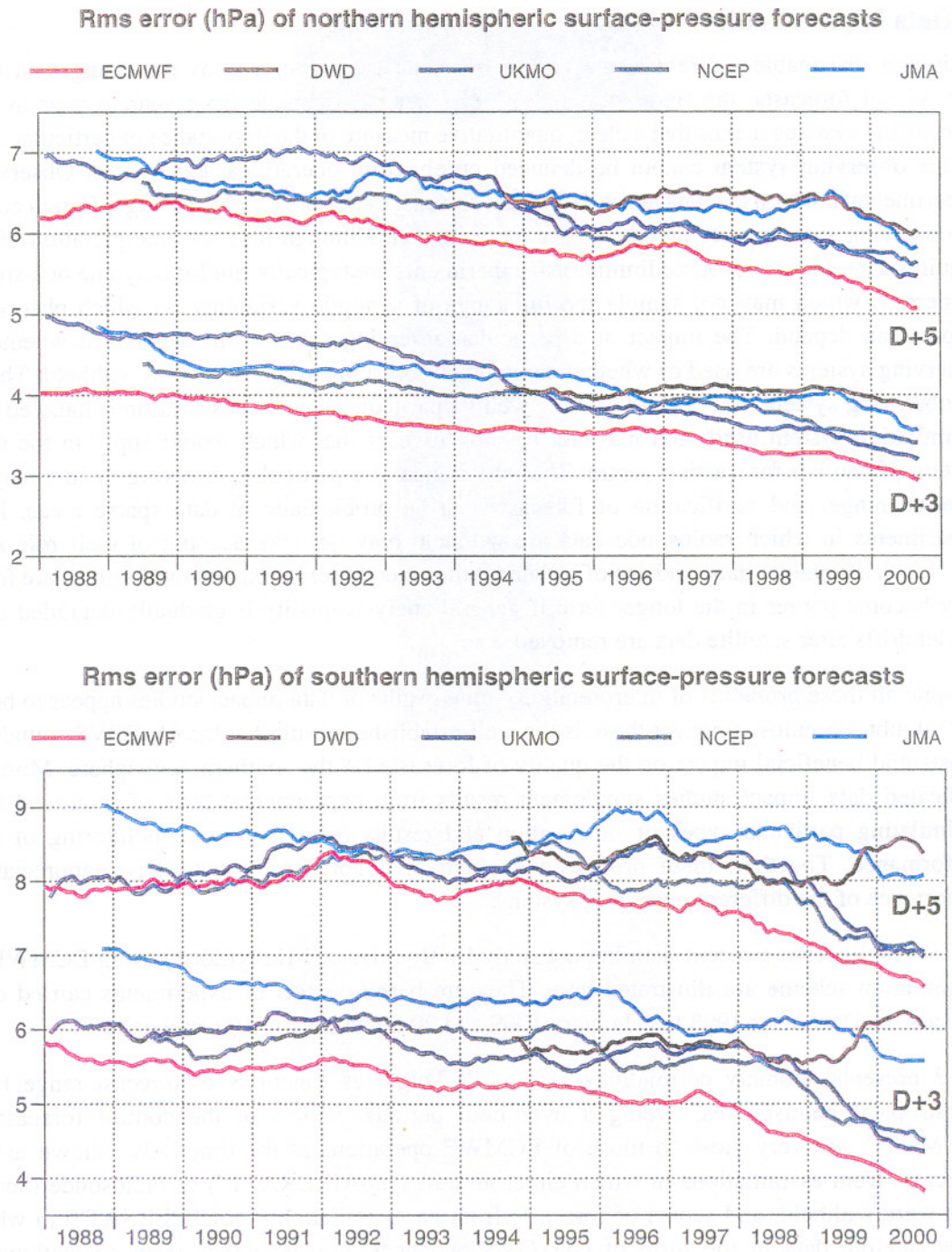

Fig. 2 Root-mean-square errors of 3- and 5-day surface-pressure forecasts for the extratropical northern and southern hemispheres produced by ECMWF, Deutscher Wetterdienst (DWD), the Met Office (UKMO), the US National Centers for Environmental Prediction (NCEP) and the Japan Meteorological Agency(JMA), plotted in the form of 12-month running means of all monthly data exchanged by the centres from July 1987 to February 2001.

McNally et al., 2000a).The southern hemispheric errors of recent DWD forecasts can be seen to be relatively large, stemming from problems in the quality of the NESDIS (A)TOVS retrievals assimilated (rather than radiances) in the DWD system (Wergen, personal communication). A more uniform behaviour is found in the northern hemisphere, although the spread in error between the forecasting centres has increased somewhat since the mid 1990s. A marked recent improvement can be seen in the forecasts from JMA, where assimilation of NESDIS ATOVS retrievals was introduced in January 2000 and 1D-Var TOVS retrievals have been used since March 2000 (Okamoto and Tada, 2001).

3. Data impact studies

Although reasonable inferences may often be made as to the reasons for changes in the skill of operational forecasts, the frequency and variety of changes made from year to year in operational forecasting systems means that a clear, quantitative measure of the importance of particular components of the observing system cannot be deduced purely from operational experience. Observing system experiments in which the data assimilation and forecast cycles are repeated using different combinations of observing system offer precise measures of impact, although here too interpretation of the results requires care. Due to resource limitations, experiments are typically run for only one or a small number of periods which may not sample the full range of synoptic variability on which observing-system impact can depend. The impact of a particular observing system will depend on whether all other observing systems are used or whether some other observing systems are also withheld. The impact of an observing system will depend on how well its particular type of observation is handled by the data assimilation system used, and may not be indicative of that which would apply in the assimilation system of another forecasting centre. The importance of a particular observing system may vary with forecast range, and verification of forecasts can be problematic in data sparse areas. In addition, experiments in which radiosonde data are withheld may not take account of their role in the bias-correction of satellite data, and use of isolated radiosonde observations (in the stratosphere for example) may become poorer in the longer term if general analysis quality is gradually degraded due to slow model drifts after satellite data are removed.

Despite all these problems of interpretation, some results of data impact studies appear to be subject to little doubt. Foremost amongst these is the well-established result that the (A)TOVS soundings have a substantial beneficial impact on the quality of forecasts for the southern hemisphere. More generally, dedicated data impact studies supplement results from experimental tests of improved methods of assimilating particular types of observation and results from long-term monitoring of operational performance. Together, these provide a qualitative, if not firmly quantitative, appreciation of the importance of the different observing systems.

A few results from a recent data-impact study by Bouttier and Kelly(2001) using ECMWF's 4D-Var assimilation scheme are illustrated here. They are based on sets of experiments carried out for two periods, 16 December 1998 to 6 January 1999 and 20 September to 10 October 1999.

Fig.3 presents anomaly correlations of 500hPa height as functions of forecast range for the two extratropical hemispheres, averaged over both periods. Scores of the control forecasts, labelled "ECMWF", are very close to those of ECMWF operations at the time. Also shown are scores of forecasts from assimilations in which either aircraft ("noAIRCRAFT") or radiosonde ("noSONDE") data were withheld, and scores of forecasts from an assimilation ("noUPPERSAT") in which upper-level satellite data in the form of (A)TOVS radiances, tracked winds from geostationary-satellite imagery and total-column water-vapour retrievals from SSM/I were withheld. Some satellite data do influence every set of forecasts, as surface wind data from the ERS2 scatterometer and SSM/I were assimilated in all cases.

The first comment to be made on Fig.3 is that removing observational data produces consistently worse forecast scores in each case, indicating that the assimilation system is behaving reasonably and that the sample size is large enough for at least broad conclusions to be drawn. Over the two periods in question denial of aircraft data results in the smallest net reduction in forecast quality, in both hemispheres. For the northern hemisphere, the detrimental impact of removing either the radiosonde data or the upper-level satellite data is similar. Conversely, by far the largest detrimental effect on southern hemisphere forecasts occurs from taking out the upper-level satellite data. Removing radiosondes has rather similar impact for the two hemispheres, and there is a little more medium-range impact in the southern than in the northern hemisphere from removing aircraft data. The impact of removing the upper-level satellite data shown here is larger than that seen by Kelly(1997) in an earlier such study (cf. his Fig. 3), and provides further evidence of recent advances in the assimilation of such data.

Fig. 4 shows corresponding verification scores indicating the impact of two of the satellite observing systems. It shows anomaly correlations for the control forecasts and for the forecasts from the

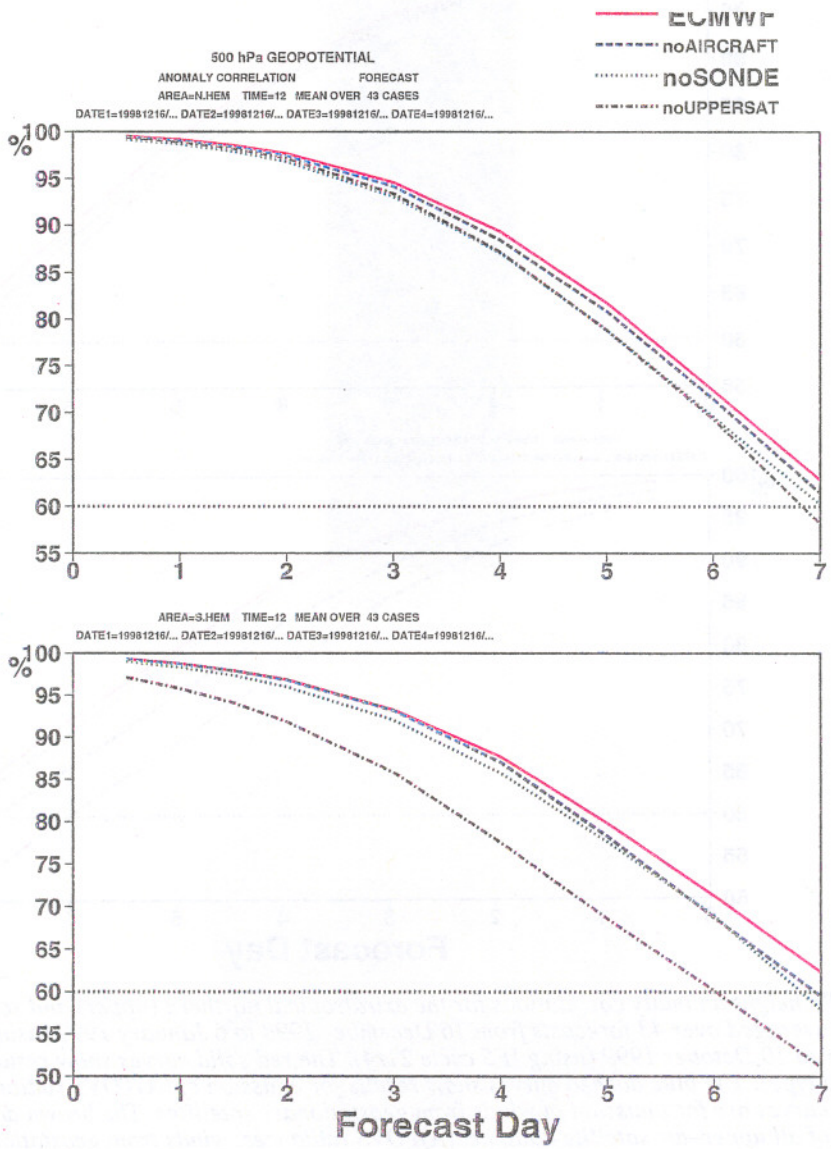

Fig. 3 500hPa height anomaly correlations for the extratropical northern (upper) and southern (lower) hemispheres averaged over 43 forecasts from 16 December 1998 to 6 January 1999 and 20 September to 10 October 1999. The red solid curves show results from assimilations using all data types. The blue dashed curves show results for omission of aircraft data and the green dotted curves are for omission of radiosonde data. The brown dash–dotted curves are for omission of upper–air satellite data, i.e. (A)TOVS radiances, winds from geostationary satellites and total–column water vapour from SSM/I.

assimilations with all upper-level satellite data withheld, as in Fig. 3, and also for forecasts from assimilations in which (A)TOVS radiances[1] ("noTOVS") and tracked winds from imagery ("noSATOB") were separately denied. There is a much larger impact of removing the radiances than of removing the tracked winds. There is also a larger impact of removing the tracked winds and SSM/I water vapour from the assimilation that already has the (A)TOVS radiances removed (seen by comparing "noUPPERSAT" with "noTOVS") than there is of removing the tracked winds alone from the complete observing system. This should not be attributed solely to the impact of the SSM/I data, as it is likely that the tracked winds play a more important role in the absence of (A)TOVS radiances than when all observing systems are in use.

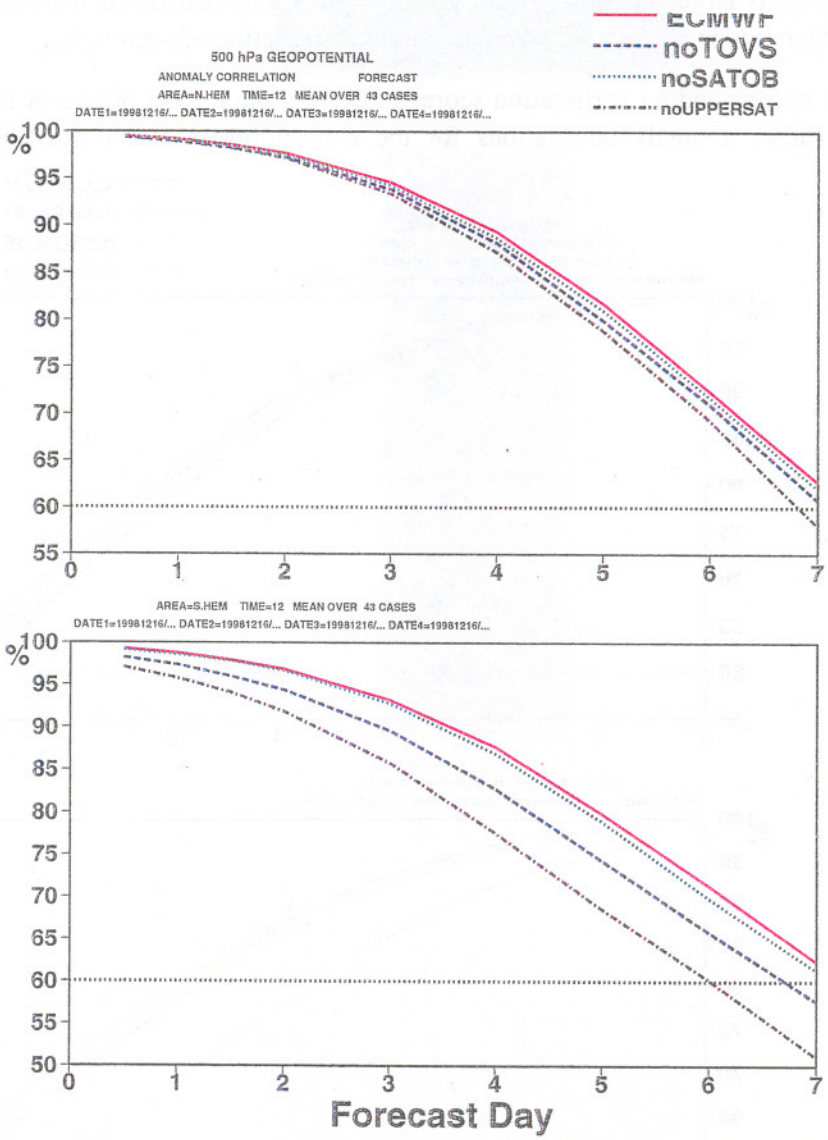

Fig. 4 500hPa height anomaly correlations for the extratropical northern (upper) and southern (lower) hemispheres averaged over 43 forecasts from 16 December 1998 to 6 January 1999 (using IFS cycle 22r1) and 20 September to 10 October 1999 (using IFS cycle 21r4). The red solid curves show results from assimilations using all data types. The blue dashed curves show results for omission of (A)TOVS radiance data and the green dotted curves are for omission of winds from geostationary satellites. The brown dash–dotted curves are for omission of all upper–air satellite data, i.e. (A)TOVS radiances, winds from geostationary satellites and total–column water vapour from SSM/I.

1. The version of the data assimilation system adopted for these experiments used only microwave radiances from the MSU and AMSU-A instruments.

A significant change has been made to the operational use of (A)TOVS radiance data at ECMWF since the above experiments were carried out. This entailed adding use of the HIRS-12 infrared water-vapour channel and the uppermost (AMSU-14) ATOVS sounding channel and relaxing some of the constraints on using other microwave channels. Following promising tests, the change in use of radiances was combined with several other changes to form a new cycle (22r3) of the ECMWF forecasting system. The other changes included a comprehensive revision of the observation and background error variances used in the data assimilation, and major changes to the model's treatment of the land surface, sea-ice and long-wave radiation. The net impact of these changes on 500hPa height anomaly correlations is illustrated in Fig. 5, which shows averages from cycle 22r3 and from the previous operational cycle (22r1) taken over a 118-day pre-operational trial period from 1 March to 26 June 2000. The forecast improvement is modest in the northern hemisphere and substantial in the southern hemisphere. Tests of the individual changes were carried out for shorter periods, and suggest that a substantial fraction (though not all) of the improvement in the southern hemisphere comes from the use of the additional radiance data.

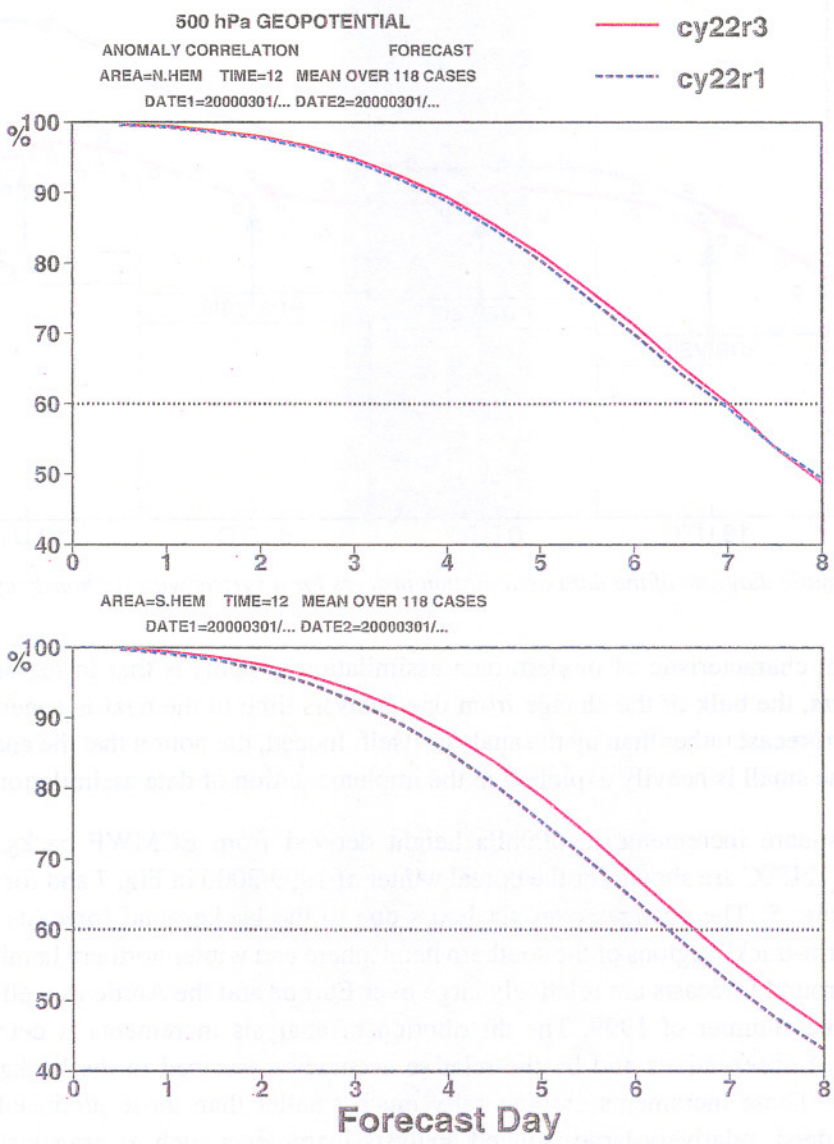

Fig. 5 500hPa height anomaly correlations for the extratropical northern (upper) and southern (lower) hemispheres averaged from 1 March to 26 June 2000 comparing results from cycles 22r3 and 22r1 of the ECMWF forecasting system.

4. Data assimilation

A sketch of the data assimilation process for an intermittent system is given in Fig. 6. In such a system, the observations (denoted by the green circles in the sketch) of a variable x within a certain time interval (six hours in the example shown) are processed to compute a correction to a *background* (or *first-guess*) forecast (or *trajectory* through phase space.) In the sketch, the correction is applied at the mid-point of the interval (the *analysis time*), as is usually the case in a three-dimensional variational or optimum interpolation analysis system. The background forecast is usually initiated at the preceding analysis time. The assimilation process produces a series of corrections (or *analysis increments*) which keep the sequence of background forecasts close to the observations, gradually moving the (discontinuous) background trajectory away from the trajectory provided by an earlier free-running forecast. The analysis at a particular time depends explicitly on the observations available for the current analysis interval, and implicitly on earlier observations through its dependence on the background forecast.

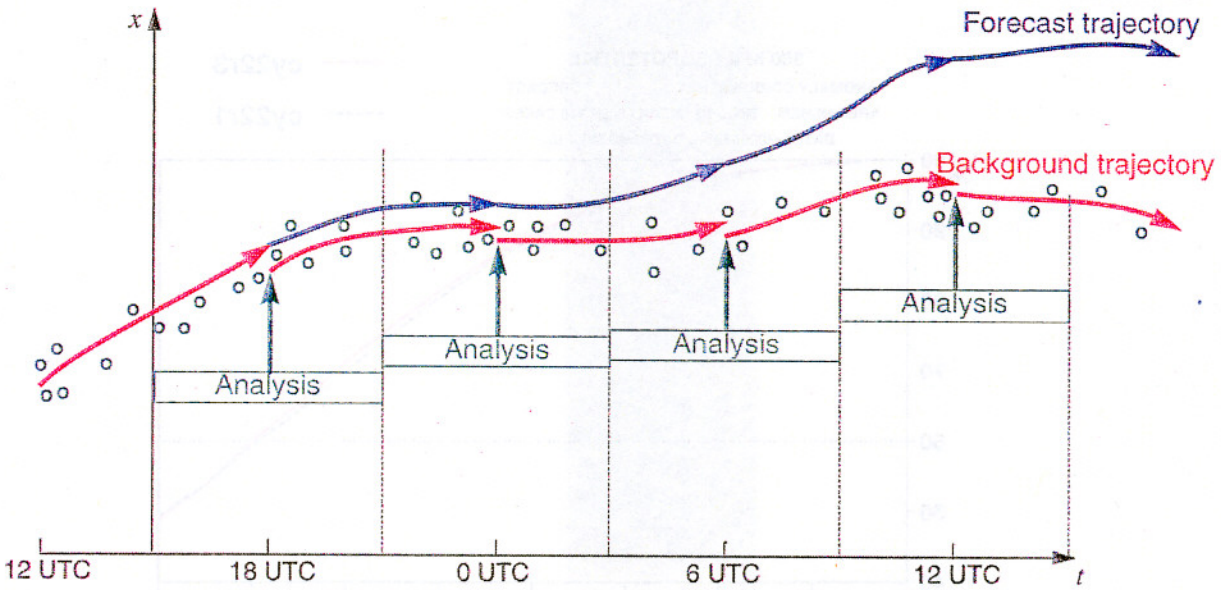

Fig. 6 Schematic diagram of the data assimilation process for a system with six-hourly cycling

An important characteristic of modern data assimilation systems is that in regions where significant change occurs, the bulk of the change from one analysis time to the next is generally provided by the background forecast rather than by the analysis itself. Indeed, the notion that the analysis increments are in some sense small is heavily exploited in the implementation of data assimilation systems.

Root-mean-square increments in 500hPa height derived from ECMWF background forecasts and analyses for 12UTC are shown for the boreal winter of 1999/2000 in Fig. 7 and for the following boreal summer in Fig. 8. The changes over six hours due to the background forecasts are largest over the oceanic "storm-track" regions of the southern hemisphere and winter northern hemisphere. Changes due to the background forecasts are relatively large over Europe and the Arctic as well as over the northern oceans for the summer of 1999. The distribution of analysis increments is determined both by the distribution of observations and by the relative accuracies ascribed to the background forecasts and observations. These increments are generally much smaller than those produced by the background forecasts. Indeed, relatively-large isolated analysis increments such as seen just north of Ellesmere Island in Fig. 7 are generally indicative either of a local problem with a particular isolated observing station (as discussed for example by Hollingsworth et al. (1986)) or of a larger-scale problem in the forecast model which is exposed only where isolated observations correct it.

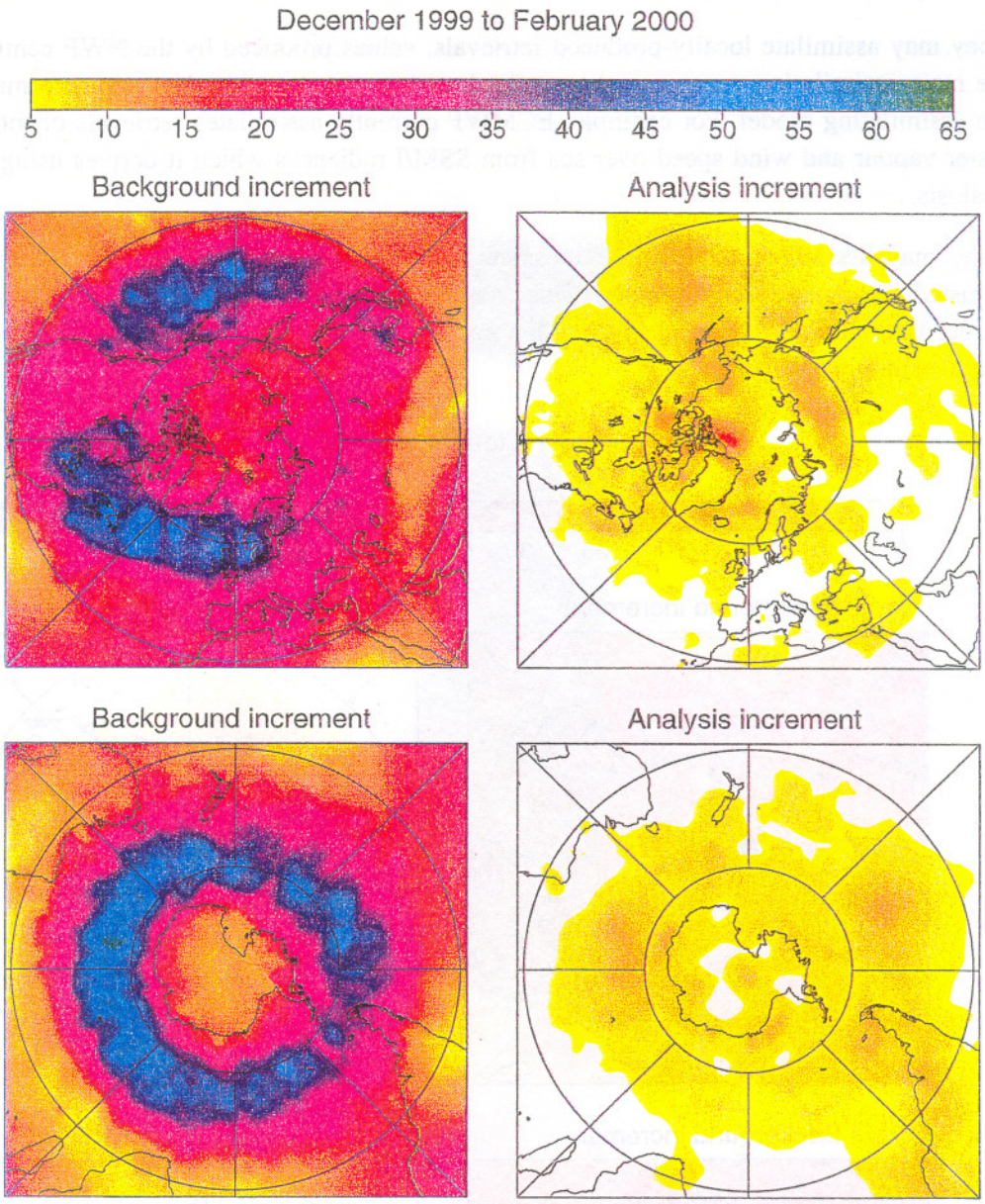

Fig. 7 The root-mean-square increment due to the background forecast from 6 to 12UTC (left), and due to the 12UTC analysis (right), for the 500hPa height field (m) computed over the period from 1 December 1999 to 29 February 2000, for the northern hemisphere (upper) and southern hemisphere (lower).

5. What is special about satellite data?

There are a number of characteristics of satellite data that make their assimilation a particular challenge.

As noted earlier, foremost among these is the fact that the measured quantities do not relate directly to the model values that are updated by the assimilation process. This results in NWP centres having three basic options for use of such data.

- They may assimilate "products" or "retrievals" disseminated by a satellite agency. These are values of conventional meteorological variables retrieved from the basic measurements made by the satellite-borne instrument. This is the original (but now widely superseded) way TOVS radiance data were utilized (as "SATEM" retrievals of thicknesses, or layer-mean temperatures).

It is still the approach universally adopted for assimilation of information from successive visible geostationary images from which cloud-track winds are derived.

- They may assimilate locally-produced retrievals, values produced by the NWP centre prior to the main assimilation process, and typically based on the use of background information from the assimilating model. For example, ECMWF currently assimilates retrievals of total column water vapour and wind speed over sea from SSM/I radiances which it derives using a 1D-Var analysis.

- They may assimilate the satellite measurements directly. In this case, model variables are adjusted during the assimilation process to produce good simulations of the measurements. Direct assimilation of the (A)TOVS radiances in a 3D- or 4D-Var data assimilation system is the prime example of this approach.

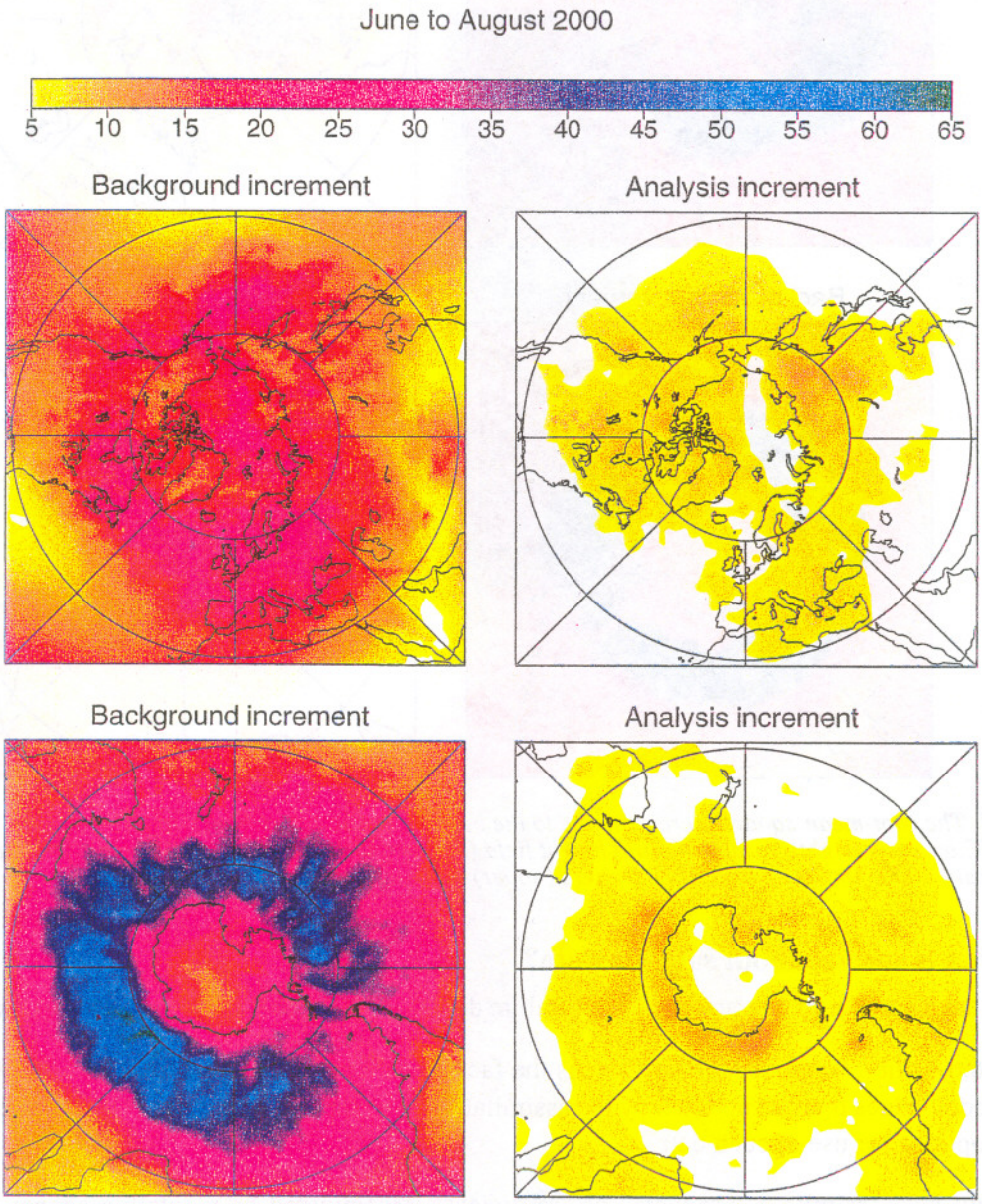

Fig. 8 The root-mean-square increment due to the background forecast from 6 to 12UTC (left), and due to the 12UTC analysis (right), for the 500hPa height field (m) computed over the period from 1 June to 31 August 2000, for the northern hemisphere (upper) and southern hemisphere (lower).

Which of these three approaches is adopted depends on a number of factors. It is clearly more straightforward to assimilate retrieved data produced by an outside agency if expertise relating to a particular instrument is lacking at an NWP centre, though the risk is that the retrieval will suffer from a poorer background (or prior) estimate. The capability of a model to simulate a particular type of observation may depend critically on the resolution of the model, horizontal or vertical, or on the realism of its treatment of clouds and precipitation. In some cases the best current approach is reasonably clear, the direct assimilation of (A)TOVS radiances for example. There are, however, also current open issues such as whether to assimilate measures of the backscattered microwave radiation, ambiguous retrieved winds or unambiguous retrieved winds from scatterometers, and whether to assimilate clear-sky water-vapour wind products or clear-sky radiances from geostationary satellites. Work towards the assimilation of radiances affected by cloud or precipitation is showing promise.

Several other characteristics either unique to satellite data or which can cause more widespread problems in the case of satellite data have also to be taken into account.

- The data are typically distributed uniformly in time rather than being available only at the main synoptic hours. This makes them particularly suited to use in a 4D-Var assimilation system.
- The data are often available with high density and correlated observation errors. They may need *thinning* or *data selection* before being supplied to the data assimilation system.
- Many of the data are difficult to use, for example due to the strong influence of cloud or the underlying surface conditions on some of the measured (A)TOVS radiances. They may thus need careful *screening* or *quality control*, which may be linked with the thinning or data selection procedure.
- Their use makes NWP vulnerable to deterioration in instrument performance or processing. The data need routine *monitoring* and occasional *blacklisting*[1] to avoid the widespread degradation of analysis fields and forecast quality that can result from erroneous data.
- The data are prone to biases. There may thus be benefit from application of *bias correction* before assimilation; this can be problematic in regions where independent data are sparse and the assimilating model also has systematic errors.

Some examples related to monitoring, quality control and bias correction will be discussed in Section 9.

6. 1D- and 3D-Variational analysis

Basic formulation

The one- and three-dimensional variational analysis problems (1D- and 3D-Var) may be stated as follows:

> Given a vector $\boldsymbol{x}_b$ of background values at time t_0, typically from a short-range forecast started from an earlier analysis, and a vector $\boldsymbol{y}$ of observations, assumed to be valid at time t_0, determine a vector $\boldsymbol{x}$ of analysis values at time t_0 that minimizes a scalar *cost function* $J(\boldsymbol{x})$, where

$$J(\boldsymbol{x}) = J_b + J_o + J_c$$

1. This typically involves *a priori* suppression from the analysis, but continued monitoring to detect when data quality improves.

with J_b measuring the difference between the analysis $\boldsymbol{x}$ and the background $\boldsymbol{x}_b$, J_o measuring the consistency between the analysis $\boldsymbol{x}$ and the observations $\boldsymbol{y}$, and J_c representing any additional constraints.

In 1D-Var, the observations $\boldsymbol{y}$ are from one instrument or a linked set of instruments (e.g. TOVS) at one point, $\boldsymbol{x}_b$ comprises a column (vertical profile) of relevant model variables (e.g. T, q and O_3) at that point, typically from the background forecast, and $\boldsymbol{x}$ provides the retrieved column of "observed" equivalents of the model variables for supply to the data analysis along with other observations.

In 3D-Var, the observations $\boldsymbol{y}$ are the complete set to be assimilated for time t_0, $\boldsymbol{x}_b$ is a three-dimensional field of all analysed model variables at time t_0, and $\boldsymbol{x}$ is the analysis of model variables at time t_0.

The definitions of J_b and J_o are generally based on quadratic forms:

$$J_b = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_b)^T\mathbf{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}_b)$$

and

$$J_o = \frac{1}{2}(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x}))^T\mathbf{R}^{-1}(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x}))$$

Here $\boldsymbol{H}(\boldsymbol{x})$ maps model variables to observed variables (e.g. T, q and O_3 profiles to simulated radiances from a particular HIRS channel), $\mathbf{B}$ is the covariance matrix of background errors, and $\mathbf{R}$ is the covariance matrix of "observation" errors.

It is important to note that $\mathbf{R}$ represents not only the effects of measurement errors, but also of errors in the *observation operator* (or *forward model*) $\boldsymbol{H}$. $\mathbf{R}$ also includes any effects of unrepresentativity arising from differences in resolution between the measurements and the model variables. It should also be noted that J_o is the sum of the separate contributions J_{oi} from each of the observations y_i if the errors of these observations are uncorrelated.

The additional constraint on the analysis represented by the J_c component of the cost function is generally used in 3D-Var (or indeed 4D-Var) to control the level of gravity-wave activity (or high-frequency oscillations) in the subsequent forecast, through formulations based on the concepts of either normal-mode initialization or digital filtering. Physical constraints could also be included, for example that the relative humidity be between 0 and 100%.

Special cases

The simplest case of variational assimilation concerns a single observation, y, of an analysis variable, x, and a single background value, x_b. For example, x may be the temperature of a room, y a current measurement with an inaccurate thermometer and x_b a measurement made several hours ago with an accurate thermometer. We denote the standard deviation of the expected error of the observation y by σ_o (the typical error of the inaccurate thermometer, which is assumed to be unbiased) and the expected error of the background value by σ_b (the typical error of assuming no change of temperature since the accurate measurement was made).

The cost function J in this case is given by:

$$J = \frac{1}{2}\left(\frac{(x-x_b)^2}{\sigma_b^2} + \frac{(x-y)^2}{\sigma_o^2}\right)$$

and the value of x which makes J a minimum is given by

$$x - x_b = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}(y - x_b) = \left(\frac{1}{1 + \sigma_o^2/\sigma_b^2}\right)(y - x_b)$$

The background value x_b is incremented by a fraction of the *innovation vector*, the difference $(y - x_b)$ between the observed value and the background value. The weighting factor depends on the ratio of the expected accuracies of the observed and background values.

A generalization of the above form can be written down in the case of small analysis increments. The observation-equivalent of the analysis increment $\mathbf{H}(x - x_b)$ can be written in terms of a weighted value of the innovation vector $(y - H(x_b))$ as follows:

$$\mathbf{H}(x - x_b) = \mathbf{H}\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(y - H(x_b))$$

where $\mathbf{H}$ is the linearized observation operator:

$$H(x) \approx H(x_b) + \mathbf{H}x - \mathbf{H}x_b \text{ for small } (x - x_b).$$

$\mathbf{H}\mathbf{B}\mathbf{H}^T$ is the covariance of background errors as represented in observation space. The analysis itself is given by

$$x = x_b + \mathbf{K}(y - H(x_b))$$

where the *gain* or *weight* matrix $\mathbf{K}$ is given by

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

Although these relationships can be simply written down, in practice they cannot be directly evaluated as the matrices involved are intractably large for full-sized NWP systems. Approximations may nevertheless be useful. For example, estimates of the diagonal elements of $\mathbf{H}\mathbf{B}\mathbf{H}^T$ provide estimates of the variances of background error in observation space, which when compared with the observation error variances provide an indication of the weight given to the observations by the analysis. Andersson et al.(2000) discuss how such estimates are made for radiance data in the ECMWF assimilation system, in which the derived background error variances in observation space are used in the quality control of observations and in diagnosis of the performance of the system.

Asynoptic observations in 3D-Var

In implementing an intermittent analysis system such as 3D-Var, a decision has to be taken as to how to use observations valid outside the main analysis times, as is particularly the case with satellite data.

Consider an assimilation cycle with period T. The analysis for time t_0 will typically use observations valid at times in a range such as $t_0 - T/2 \le t < t_0 + T/2$. The basic approach adopted in ECMWF's former operational implementation of 3D-Var was to use only observations made at the closest times to t_0 from fixed or slowly moving platforms, and to use all observations from fast moving platforms, satellites in particular. All observations used were assumed to be valid at t_0.

An improved approach known as FGAT (First Guess at the Appropriate Time) was implemented in 1991 in ECMWF's then-operational optimal interpolation analysis. It has recently been adopted for the ERA-40 re-analysis project, which is using the ECMWF 3D-Var assimilation system to produce a set of analyses for the period from mid-1957 to present. It is also being used for the six-hourly short-cutoff 3D-Var analysis recently implemented at ECMWF from which a forecast is run to provide boundary conditions for Member-States' short-range limited-area forecast models. Such an approach was also

introduced in the Met Office's 3D-Var assimilation scheme in May 2000 (Lorenc et al., 2000; Bell et al., 2000). In the FGAT approach, the deviation of the background from the observation (the innovation vector) is evaluated at (or nearer to) the actual observation time, but it is utilized in 3D-Var as if it is valid at the analysis time t_0. Expressed mathematically, the term $y_i - H_i(\boldsymbol{x})$ that enters the contribution to the observation cost function component J_o from observation y_i is replaced by $(y_i - H_i(\boldsymbol{x}) + H_i(\boldsymbol{x}_b) - H_i(\boldsymbol{x}_{bi}))$, where $\boldsymbol{x}_{bi}$ is the vector of model variables from the background forecast valid at (or nearer to) the observation time t_i.

7. 4D-Var

The basic process

A sketch illustrating a cycle of the four-dimensional variational (4D-Var) data assimilation process is given in Fig. 9. The specific case is one of 12-hourly cycling as run operationally at ECMWF since September 2000. In this example the background forecast (or initial trajectory) is run from the preceding analysis (nominally valid for 0UTC) and the observations to be assimilated are distributed from 3UTC to 15UTC. The 4D-Var analysis seeks to determine iteratively the correction to the background state at the start of the assimilation window (3UTC in this case) that ensures the best fit of a (12-hour) forecast (or model trajectory) to the observations available over the assimilation window, taking into account the expected error of the observations and background forecast. The optimally-corrected (final) 12-hour trajectory can be simply extended in time to provide the actual forecast, although in the ECMWF system surface fields from the final trajectory are adjusted at the nominal analysis time of 12UTC based on separate surface analyses. The daily medium-range forecast is then run from the resulting adjusted model state for 12UTC. This forecast is also formally the background forecast for the analysis of observations for the subsequent assimilation window, which runs from 15UTC to 3UTC. In practice however, the forecast has to be repeated later, following receipt of the observations. This is because comparison of the background forecast with observations is needed both for screening observations to reject those that differ substantially from the background, and to measure the fit of the background to the accepted observations needed to begin the iterative process to determine the next analysis.

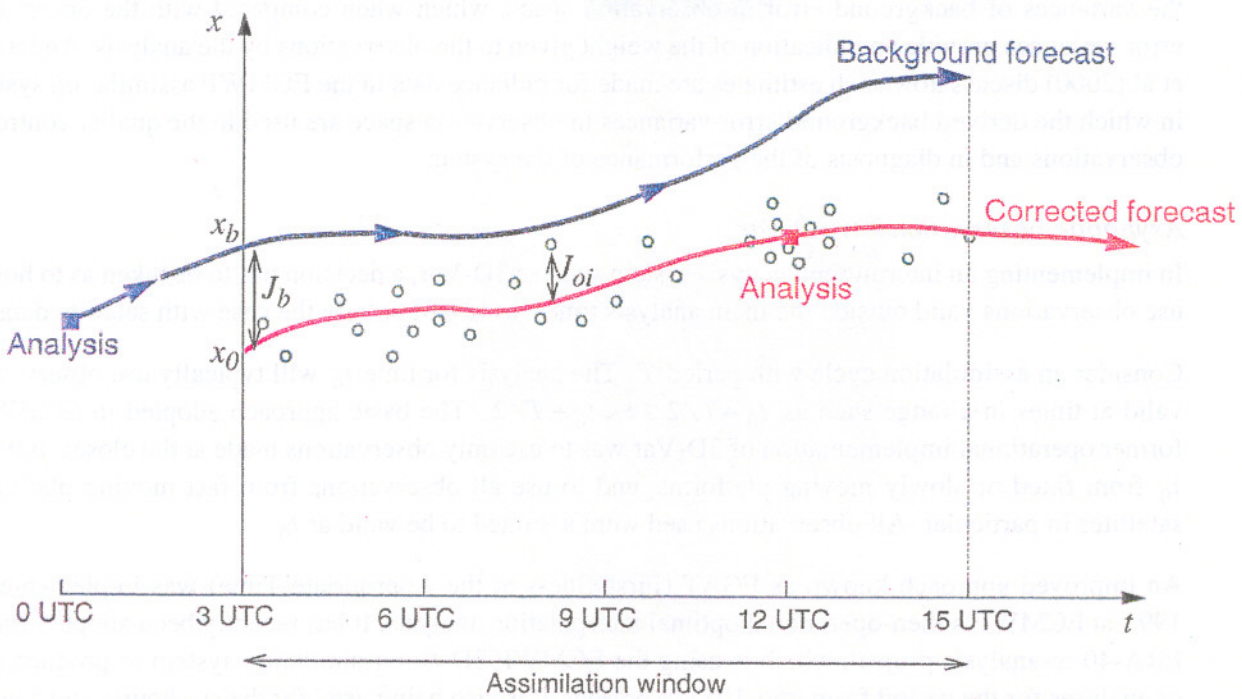

Fig. 9 Schematic diagram of 4D-Var data assimilation for a 12-hour cycle as implemented at ECMWF

The analysis problem

In 4D-Var, the observational component of the cost function, J_o uses observations $\boldsymbol{y}_i$ and model values $\boldsymbol{x}_i$ at times t_i in the range $t_0 \le t < t_0 + t_N$, with the $\boldsymbol{x}_i$ taken from a forecast starting at t_0. J_o is defined by:

$$J_o = \sum_{i=0}^{i=N} \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{H}(\boldsymbol{x}_i))^T \mathbf{R}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{H}(\boldsymbol{x}_i)) = \sum_{i=0}^{i=N} J_{oi}$$

J_b has the same form as in 3D-Var, but is evaluated using values at the beginning of the assimilation window, at time t_0. The analysis problem in 4D-Var is to determine $\boldsymbol{x}_0$, the model state at t_0, that minimizes $J_b + J_o + J_c$.

The problem is solved iteratively using a minimization procedure which utilizes one or more pairs of values of the cost function J and its gradient with respect to the analysis state, $\nabla_{x_0} J$, to refine the estimate of $\boldsymbol{x}_0$ at each iteration. The main computational task is to evaluate J_o and its gradient. This entails two steps for each pair of values. As explained by Talagrand and Courtier(1987), the first is forward integration of the forecast model from t_0 to t_N, storing the $\boldsymbol{x}_i$ and enabling computation of J_o:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{M}(\boldsymbol{x})$$

The second is backwards integration from t_N to t_0 of an inhomogeneous form of the adjoint of the tangent-linear version of the model, the version linearized about the forward trajectory $\boldsymbol{x}_0$ ... $\boldsymbol{x}_N$. The forcing term included in the inhomogeneous adjoint model depends on the departures $\boldsymbol{y}_i - \boldsymbol{H}(\boldsymbol{x}_i)$ between observations and their equivalents derived from the forward trajectory:

$$\frac{d}{dt}\boldsymbol{x}^* = -\mathbf{M}^*\boldsymbol{x}^* - \mathbf{H}^*(\nabla_H J_{oi})\delta(t - t_i)$$

The variable $\boldsymbol{x}^*$ is initially set to zero ($\boldsymbol{x}^*_N = 0$) and the solution $\boldsymbol{x}^*_0$ of the adjoint integration at time t_0 is the required gradient $\nabla_{x_0} J_o$.

Incremental 4D-Var

If the basic 4D-Var method outlined above requires 100 evaluations of the cost function and its gradient to achieve an acceptable minimisation, and if the adjoint model requires twice as much computation as the forward model, then the model-integration component of processing observations for a 24-hour period using 4D-Var is as demanding as producing a standard 300-day forecast. The "incremental" formulation of 4D-Var was proposed by Courtier et al. (1994) as a way of reducing the computational cost and thereby making operational implementation of the method feasible. The cost is reduced by carrying out the minimisation using a model with lower resolution and simplified physical parametrizations.

The incremental approach utilizes a linearized form of the observation cost function given by

$$J_o = \sum_{i=0}^{i=N} \frac{1}{2}(\boldsymbol{d}_i - \mathbf{H}\delta\boldsymbol{x}_i)^T \mathbf{R}_i^{-1} (\boldsymbol{d}_i - \mathbf{H}\delta\boldsymbol{x}_i)$$

where $\delta\boldsymbol{x}_i = \boldsymbol{x}_i - \boldsymbol{x}_{bi}$, $\boldsymbol{d}_i = \boldsymbol{y}_i - \boldsymbol{H}(\boldsymbol{x}_{bi})$ and $\boldsymbol{x}_{bi}$ is result of running the full model from t_0 to t_i with initial state $\boldsymbol{x}_b$. The procedure entails solving the minimization problem for a lower-resolution representation of $\delta\boldsymbol{x}_0$, using a forward run of a lower-resolution tangent-linear model with simplified parametrizations and a backward run of the adjoint of this model forced by observation departures $\boldsymbol{d}_i$. The lower-resolution increment $\delta\boldsymbol{x}_0$ is then interpolated to the higher resolution of the full model and added to $\boldsymbol{x}_b$. The procedure may then be repeated one or more times, replacing $\boldsymbol{x}_{bi}$ by the result of

integrating the full model from t_0 to t_i with initial state $\boldsymbol{x}_b + \delta\boldsymbol{x}_0$. The resolution or complexity of the parametrizations of the minimizing linear model may be increased in any of these later minimizations.

Configurations for global operational 4D-Var data assimilation have been developed by ECMWF and Météo-France, both utilizing a common, jointly-developed forecasting-system code (ARPEGE/IFS). In the first operational version implemented at ECMWF in November 1997 (Mahfouf and Rabier, 2000), the horizontal resolution of the full model used in the outer loops of the incremental procedure was T213 (corresponding approximately to a 90km half-wavelength for the smallest-resolved scale, with a grid resolution of about 60km). The lower resolution used in the minimising inner loops of the procedure was T63 (with 320km smallest half-wavelength and 210 km grid). Two outer loops were used, with more comprehensive parametrizations in the second minimization. The semi-Lagrangian advection scheme used in the full model enabled the spectral resolution of this model to be increased to T319 in April 1998, the grid resolution being unchanged. The cycling period was initially six hours, but was increased to 12 hours in September 2000. This was followed by a major upgrade in November 2000 in which the full-model resolution was increased to T511 (40 km grid) and the resolution of the (now semi-Lagrangian) tangent-linear and adjoint models was increased to T159 (130km). A further increase in inner-loop resolution to T255 (80km) is under development.

Météo-France implemented a version of 4D-Var in operations in June 2000 in which the inner-loop resolution was progressively increased with each outer iteration of the incremental procedure (Thépaut, personal communication). Three outer loops were used, with resolutions of T42, T63 and T95 for the (Eulerian) inner loops in the three minimizations, each of which used 25 iterations (to be compared with the 50 and 25 iterations used in the two minimizations in the current ECMWF version). As in the ECMWF version, comprehensive parametrizations were used only in the final 25-iteration minimization. The outer-loop resolution of the model was T199, but with coordinate stretching (Courtier and Geleyn, 1988) by a factor 3.5 resulting in a local resolution equivalent to around T700 over France.

8. Further Remarks

It is not possible to give here an exhaustive account of additional properties, assumptions and possible extensions of variational data assimilation, but it is appropriate to mention briefly some items which are of particular relevance to the assimilation of satellite data.

- The variational analysis may be extended to determine parameters or variables in addition to the initial model state variables which are the primary target of the process. Examples are the determination of radiance bias-correction coefficients (implemented at NCEP as discussed by Derber in these Proceedings) and of radiative skin temperature at satellite measurement points (implemented at ECMWF for assimilation of raw radiances).

- The formulation as set out here, and as generally implemented, makes no allowance for correlation of background and observation errors. Some degree of correlation may, however, exist due either to assimilating background-dependent 1D-Var retrievals in a 3D- or 4D-Var system, or to utilizing background values in observation quality control or selection procedures.

- Implementation of horizontal correlation of the errors of observations from a particular satellite system is not entirely straightforward, but is desirable in principle. It should be recalled that in this context "observation" errors means not only measurement errors but also errors in the observation operator, for example the fast radiative transfer model used to calculate background radiances, which may be subject to errors which depend on airmass characteristics. A practical alternative to introduction of horizontal correlation of observation errors is to thin the observations more than would otherwise be done.

- The quadratic form of J_o is appropriate for a distribution of observation errors that is Gaussian in form. The assimilation of ERS scatterometer data at ECMWF uses a different form, based on combining two separate cost functions of quadratic form, each computed for one of the pair of ambiguous retrieved winds that is presented to the analysis (Stoffelen and Anderson, 1997).

- Variational quality control simply reduces the influence of a particular observation for uncorrelated observation errors. The reduction depends on a derived probability of gross error, which modifies the quadratic form of J_o used for most observations (Andersson and Järvinen, 1999).

- A strictly quadratic cost function may be needed for some applications. In particular, this is a requirement of the method used at ECMWF to estimate the spatial distribution of analysis error variance (Fisher, 1996). The estimation (in contrast to the 4D-Var analysis itself) consequently uses no variational quality control and a modified treatment of scatterometer data.

9. Monitoring, quality control and bias correction

Routine monitoring of the observing system is an important function of a comprehensive data assimilation system used for operational forecasting or re-analysis. This is particularly important in the case of satellite data since, as noted earlier, deterioration in instrument performance or a problem in data processing can rapidly cause widespread degradation of analysis and forecast quality. An example of this is provided by comparison of results from ECMWF's two re-analysis projects, the original ERA-15 project analysing the period from 1979 to 1993 and the new ERA-40 project. Fig. 10 shows ERA-40 monitoring plots comparing time series of global-mean 200hPa temperature from a pre-production version of the ERA-40 analyses and from the ERA-15 production analyses, for the period from 1 September to mid-November 1986. For the first two months there is quite a close agreement between

Fig. 10 The global mean (red) and the standard deviation (green) of differences between ERA-40 and ERA-15 analyses (upper), and the global means of ERA-40 (blue) and ERA-15 (red) analyses (lower), for 200hPa temperature (°C), plotted for each analysis cycle from 1 September to 15 November 1986.

the mean temperatures from the two analyses, ERA-40 tending to be slightly warmer and less variable from one cycle to the next. Then, on 2 November the ERA-15 analysis cools substantially due to a sudden degradation in quality of MSU-3 data from NOAA-9 following a cosmic storm. This problem was not detected in time to enable remedial action to be taken during ERA-15 production, and it had a persistent detrimental effect due to an impact on bias correction that lasted beyond the period of assimilation of degraded data (Fiorino, 2000). It was avoided in the ERA-40 analyses by the expedient of blacklisting the corrupted data from 24 October onwards.

Corresponding global maps of the differences in 200hPa temperature between the pre-production ERA-40 analyses and the ERA-15 analyses are presented in Fig. 11 for 1, 2, 3 and 5 November 1986. On 1 November, before the corruption of the satellite data, differences are mostly located in the southern hemisphere, and neither positive nor negative values predominate. By 5 November however, the ERA-15 analyses have become several degrees cooler than the ERA-40 analyses over much of the southern hemisphere, Tropics and North Pacific. Differences remain small over most of the rest of the northern hemisphere, where there are sufficient alternative data to control the near-tropopause temperature analyses. There is a persistent difference over China because biases in radiosonde temperatures were corrected in the ERA-15 assimilation, whereas they were not in this particular pre-production ERA-40 assimilation, which in fact was being run for a year to generate the statistics necessary for subsequent radiosonde bias correction. This serves as a reminder that it is not only the satellite-based component of the observing system that requires monitoring and possible action to adjust for biases or to blacklist when appropriate.

It is desirable that problems in the incoming data can be detected automatically by the assimilation system, and that seriously inaccurate data can be excluded from the assimilation without the need for manual addition to a blacklist. This is particularly important when the data are subject to relatively frequent, short-lived inaccuracies. An example of this is provided by the scatterometer data from the ERS-2 satellite. Periodic orbital-correction manoeuvres of the satellite degrade scatterometer data quality substantially, especially measurements made for large scan angles (measurements with high "node numbers"). In this case most of the suspect data can be detected and eliminated from the analysis by a consistency check based solely on the data received in a particular time interval. The backscatter measurements from the three (fore-, mid- and aft-) beams of the instrument are checked for their closeness to a conical distribution in the three-dimensional measurement space (discussed by Stoffelen in these Proceedings), and all data for the time interval are excluded once the mean "distance to the cone" exceeds a critical value. The solid lines in the upper set of panels in Fig. 12 show time series of the mean normalised distance to the cone for different sets of node numbers from 20 August to 5 September 1999. During this period there were two instances (on 25 August and 1 September) of abnormally large distances to the cone for high node numbers. The lower set of panels shows that at these times there were abnormally large biases of retrieved winds compared to background winds, evident for all node numbers, providing additional justification for rejection of the observations.

An example of a more selective pre-screening of observations before assimilation is provided by the way SSM/I retrievals are currently used in ECMWF's assimilation system. The retrieval scheme for these data utilizes 1D-Var with a microwave radiative transfer model which makes no allowance for scattering by raindrops and ice particles. Consequently, the retrievals are relatively inaccurate in areas where there is rainfall. Retrievals are thus selected for assimilation only at points which are not flagged as raining by the SSM/I-based rainfall-detection algorithm of Bauer and Schlüssel (1993). An illustration is given in Fig. 13, which is adapted from work of Marécal and Mahfouf (2000) that led to the operational implementation of this pre-screening in April, 2000. It is referred to as pre-screening to distinguish it from the screening carried out within the assimilation process that rejects any observation which departs substantially from its background value.

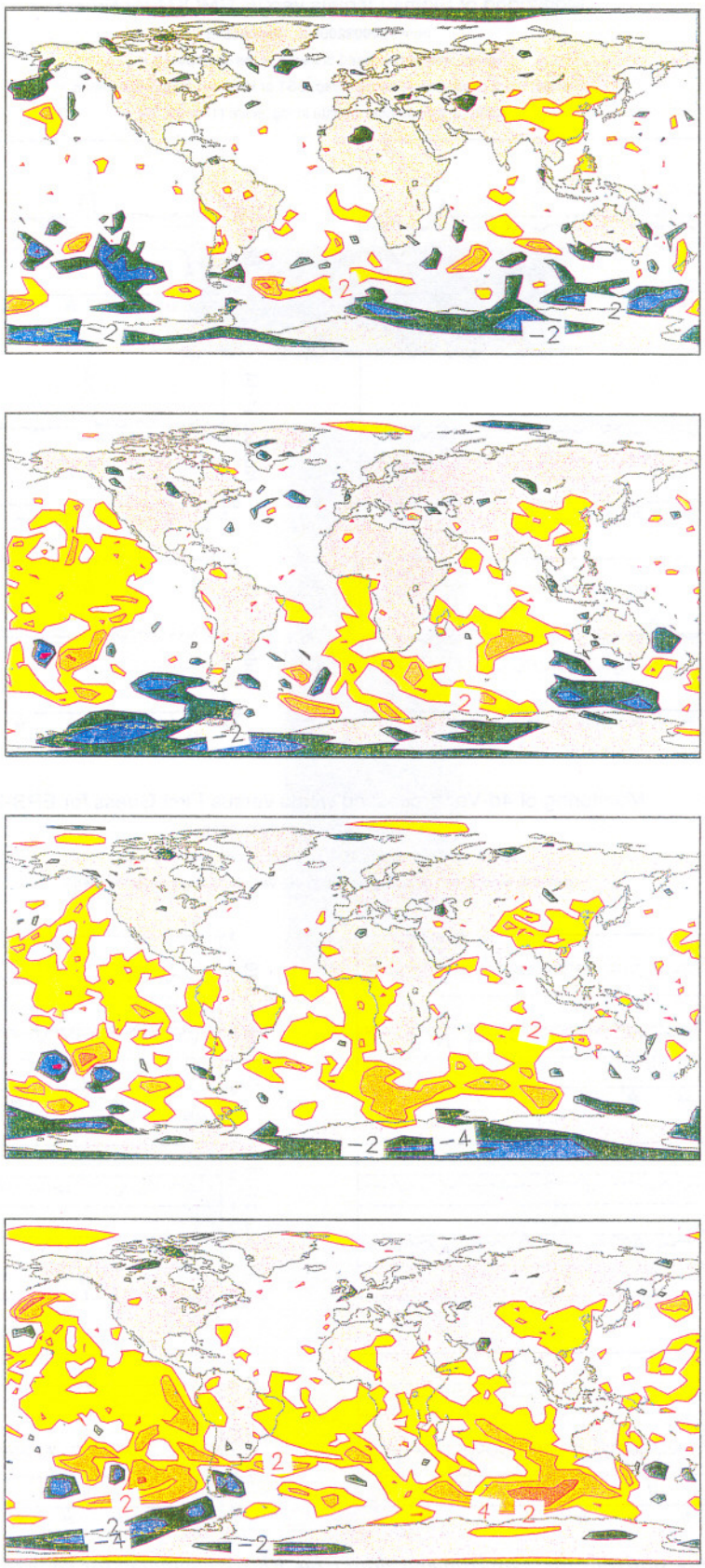

Fig. 11 Difference between ERA-40 and ERA-15 analyses of 200hPa temperature for 12UTC on 1, 2, 3 and 5 November 1986. Contours are ±2K, ±4K and ±8K. Yellow and orange shading indicate positive values (i.e. ERA-warmer than ERA-15). Green and blue shading indicate negative values.

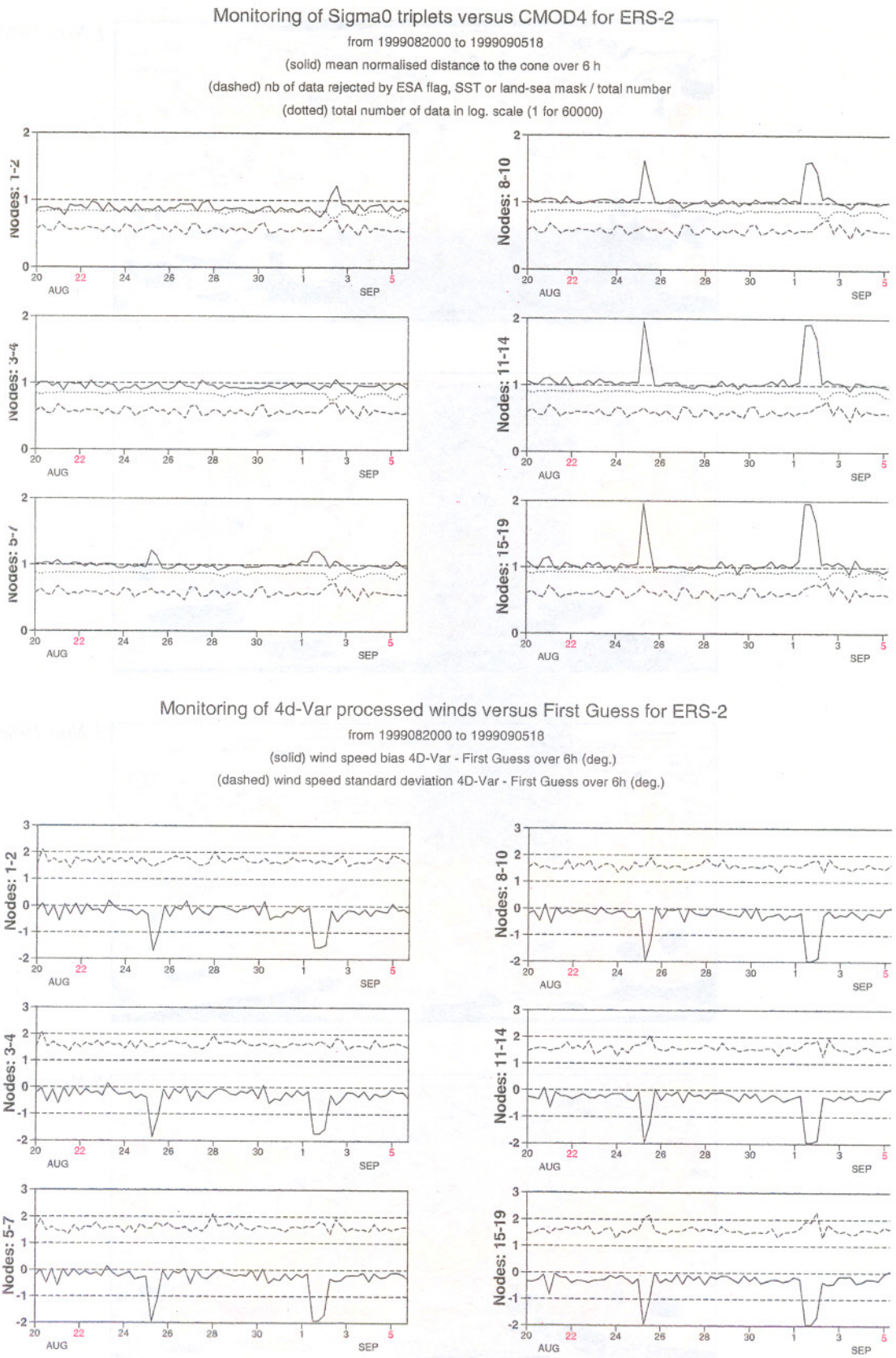

Fig. 12 Monitoring statistics for scatterometer data from ERS-2 computed for each 6-hour analysis interval from 20 August to 5 September 1999.The upper panels shows consistency checks based solely on the backscatter measurements and the lower plots shows statistics of differences between retrieved wind observations and the background winds.

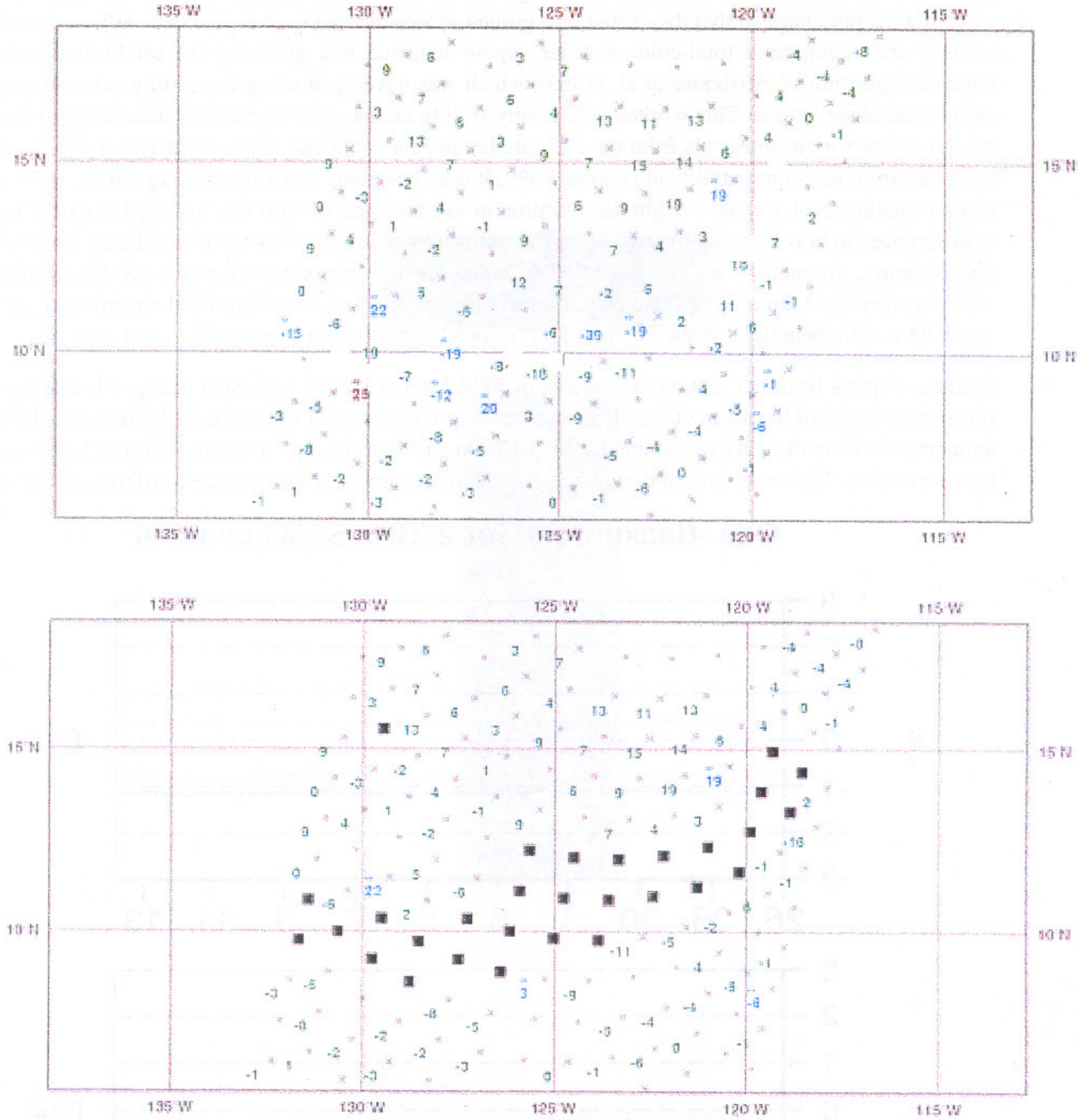

Fig. 13 Differences between wind speeds retrieved from SSM/I observations using 1D-Var and background wind speeds, plotted as values at observation points in units of 0.1ms^{-1}, for 12UTC 7 September 1999. The upper panel shows all values that would be used by the data assimilation in the absence of a rainfall check. The lower panel shows the values used when retrievals are suppressed at the points (denoted by black squares) where rainfall is detected by the algorithm of Bauer and Schlüssel(1993).

Bias-correction is an important component of a complete satellite data assimilation system. The scheme used to correct biases in (A)TOVS data at ECMWF is described by Harris and Kelly (2001). It comprises two parts. The first is a scan-dependent correction, which applies a latitudinally varying adjustment to radiances to remove the mean dependence of background (or, as more recently applied, analysis) departures on scan position. The second is an airmass-dependent correction, which applies an adjustment based on linear regression using airmass predictors based on background (or analysed) fields, with regression coefficients determined using a set of departure statistics accumulated in the vicinity of radiosonde locations.

Bias-correction of the SSM/I radiances was originally done in a quite different manner (Gérard, 1999), adjusting the radiances so that they fitted background values that themselves had been adjusted so that globally the background total-column water vapour matched that given by the established SSM/I retrieval algorithm of Alishouse et al. (1990), which was developed using a carefully selected set of radiosonde observations. This worked sufficiently well to enable operational implementation of total-column water vapour retrievals from the F13 satellite in June 1998 and implementation of wind-speed retrievals from the same satellite in October 1999. It was, however, found that adding retrievals from a second satellite, F14, caused a slight deterioration in forecast quality. This was ascribed in part at least to differences in bias between the two satellites that were not corrected by the method then employed. Subsequently, the method applied to (A)TOVS radiances has been extended to the SSM/I radiances. This plus increased thinning enabled data from F14 as well as F13 to be assimilated operationally from April 2000, with beneficial impact on low-level tropical and southern hemispheric wind forecasts.

Extracts of plots from the routine monitoring of SSM/I radiances are presented in Fig. 14. The figure shows time series of the mean of the difference between observed and simulated background brightness temperatures from the 22GHz channel for the F13 and F14 satellites, both before (blue) and after (red) bias correction. The mean plus and minus one standard deviation of the corrected differences are also

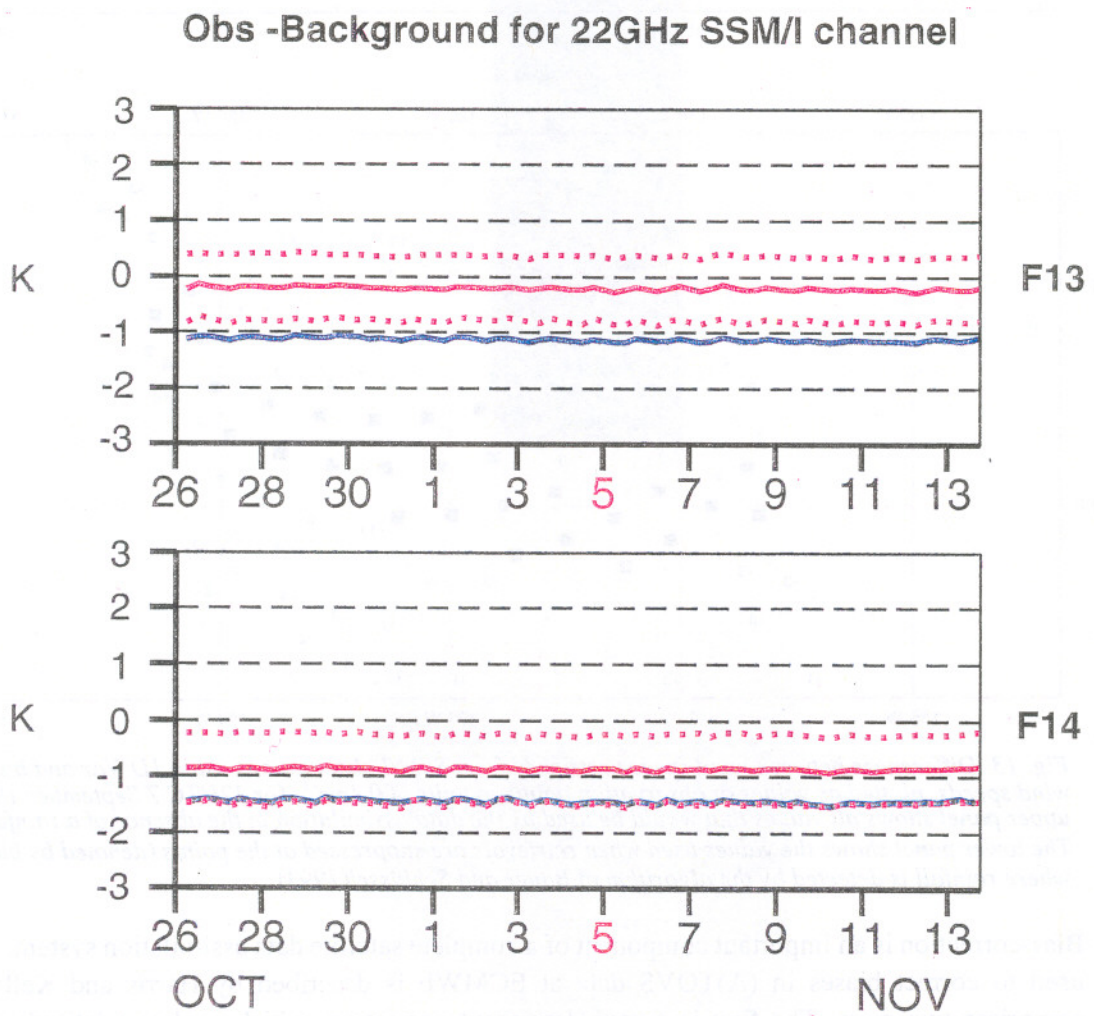

Fig. 14 Differences between observed and background values of brightness temperature (K) for the 22GHz channel of the SSM/I instruments on the F13 (upper) and F14 (lower) DMSP satellites, averaged over all observations processed for each assimilation cycle from 26 October to 13 November 2000. The solid blue line shows the mean differences for the uncorrected observations and the solid red line the mean differences for bias-corrected observations. The dotted red lines show the mean plus and minus one standard deviation of the differences for corrected observations.

plotted. The sequence shows the bias-correction scheme to be working as intended for F13, with little mean difference between the background and corrected radiances. During the period illustrated, however, the correction applied to the F14 data was inadequate, the corrected brightness temperatures being on average almost 1K colder than the values computed from background fields. This arose because of an earlier shift in the radiances received from F14, and necessitated a subsequent recalculation of the bias-correction coefficients for this satellite. This could have been avoided by using a running recalculation of the coefficients each analysis cycle, perhaps as part of the variational analysis itself as noted earlier. In general, however, the biases from the assimilated ATOVS and SSM/I radiances are quite stable in time. A considered, intermittent recalculation of coefficients currently appears to be acceptable operationally for these radiances, provided the monitoring statistics are utilized to signal promptly the occurrence of a significant deviation from the norm.

It was noted earlier that bias correction can be problematic in regions where independent data are sparse and the assimilating model is prone to biases also. This is particularly the case at upper, near-stratopause levels in the current 60-level version of the ECMWF system used in operations and for ERA-40, which has its uppermost four levels at pressures lower than 1hPa. This is shown clearly by early experience with ERA-40. In ERA-40's first production run past 24 October 1986, data from the uppermost channel of the stratospheric sounding unit, SSU-3, was inadvertently blacklisted at the same time data from MSU-3 was intentionally blacklisted to avoid use of data corrupted by the cosmic storm noted earlier. SSU-3 data in ERA-40 (and data from AMSU-A channel 14 in current ECMWF operations) play an important role in inhibiting drift in the assimilating model near the stratopause, where very little other data are available on a routine basis. The removal of the SSU-3 data in the first ERA-40 run showed up immediately in monitoring plots based on time series of the global means and standard deviations of analysis increments for 1hPa temperature. It led over a few days to a 5°C fall in mean 1hPa temperature, as illustrated in Fig. 15. The production run was subsequently repeated from 24 October onwards with data from SSU-3 reinstated, and with relatively noisy data from HIRS-1 (the uppermost HIRS channel)

Fig. 15 The global mean (red) and the standard deviation (green) of analysis increments from the first production run of ERA-40 (upper), and the global means of the corresponding analysis (blue) and background (red) fields (lower), for 1hPa temperature (°C), plotted for each analysis cycle from 1 September to 14 November 1986.

suppressed. It gave a stable trace of global mean 1hPa temperature following an initial rise of about 0.5°C associated with removing the HIRS-1 data (not shown).

When ERA-40 production is carried out for earlier years there will almost certainly be a large jump in the upper stratospheric analysis when SSU data first become available, as analyses for the pre-satellite years will be dominated by the model climatology rather than being influenced by observations at the uppermost levels. Indeed, a warming over a week or so from 1 September can be seen in Fig. 15 as ERA-40 production was spun up from a background field taken from an assimilation that used no satellite radiance data. The temporal continuity of upper stratospheric ERA-40 analyses will also be vulnerable to any gaps in coverage or drifts in bias of SSU-3 that might occur during the TOVS period.

There remains the question of what if any bias correction to apply to data from the high-level sounding channels in these circumstances. SSU-3 data in current ERA-40 production are being adjusted by a fixed offset that was determined from analyses produced during a period of concurrent operation of the SSU on NOAA-11 and AMSU-A on NOAA-15. Data from the uppermost channel of AMSU-A were used without bias correction, based on comparisons with independent satellite data (HALOE on UARS) and lidar data, which pointed to the bias between the model and AMSU channel-14 data as being primarily due to systematic model error (McNally et al., 2000b). Brightness temperatures from AMSU channel 14 on the recently launched NOAA-16 satellite were subsequently found to differ from corresponding NOAA-15 temperatures by under 1K in the mean.

Currently, the AMSU channel-14 data from NOAA-16 are assimilated in operations at ECMWF without bias correction. These data were introduced operationally a week after the failure of channel 14 on NOAA-15, during which time channel-14 brightness temperatures simulated from the background fields of the data assimilation shifted by about 2K compared with the NOAA-16 radiances, which were being monitored passively by the data assimilation system at the time. Fig. 16 presents a radiance monitoring plot for NOAA-16, showing a time series of the mean (and the mean plus and minus one standard deviation) of the difference between observed NOAA-16 and simulated background channel-14 brightness temperatures. The NOAA-16 data first became available to ECMWF on 12 October 2000, and passive monitoring began immediately. Until the last day of the month this monitoring was against an assimilation which utilized the NOAA-15 radiances. A mean difference of about 1K can be seen in this period. In the following week, almost no channel-14 data were assimilated following the failure of the channel on NOAA-15, and a sharp rise in the mean difference between the background and the passive NOAA-16 observations can be seen. NOAA-16 data were made active in the data assimilation from 8 November onwards, and the assimilation adjusted rapidly to them, brightness temperatures settling down to a mean level less than 1K colder than in the period when the NOAA-15 channel-14 data were being assimilated.

A subsequent period of poorer fit of the background to the observations peaking around 9 December can also be seen in Fig. 16. At the time this caused initial concern that there was a further instrument problem with which to contend, but it was found that the winter stratospheric vortex had become elongated and centred well off the North Pole, and the behaviour was ascribed to variability in sampling as successive NOAA-16 orbits intersected the vortex differently.

10. Conclusion

In May 1991, NESDIS TOVS retrievals for the northern hemisphere troposphere were removed from the set of observations used by the operational ECMWF data assimilation system following evidence from observing system experiments that their use had a detrimental impact on analysis and forecast quality in the region. Just over one year later, TOVS retrievals for the region were reintroduced, this time produced within the ECMWF assimilation system by the 1D-Var method and presented to the then-operational optimal interpolation analysis. This was the first step in a sequence of major improvements

(made at ECMWF and at other global forecasting centres) in the processing of radiance data, the method of analysis and the quality of the assimilating model which together with instrumental improvements and improved derivation of products such as cloud-tracked winds have established satellite-based measurements as a major component of the operational observing system for the northern hemisphere. At the same time, these developments have brought about an even more substantial increase in the quality of forecasts for the southern hemisphere.

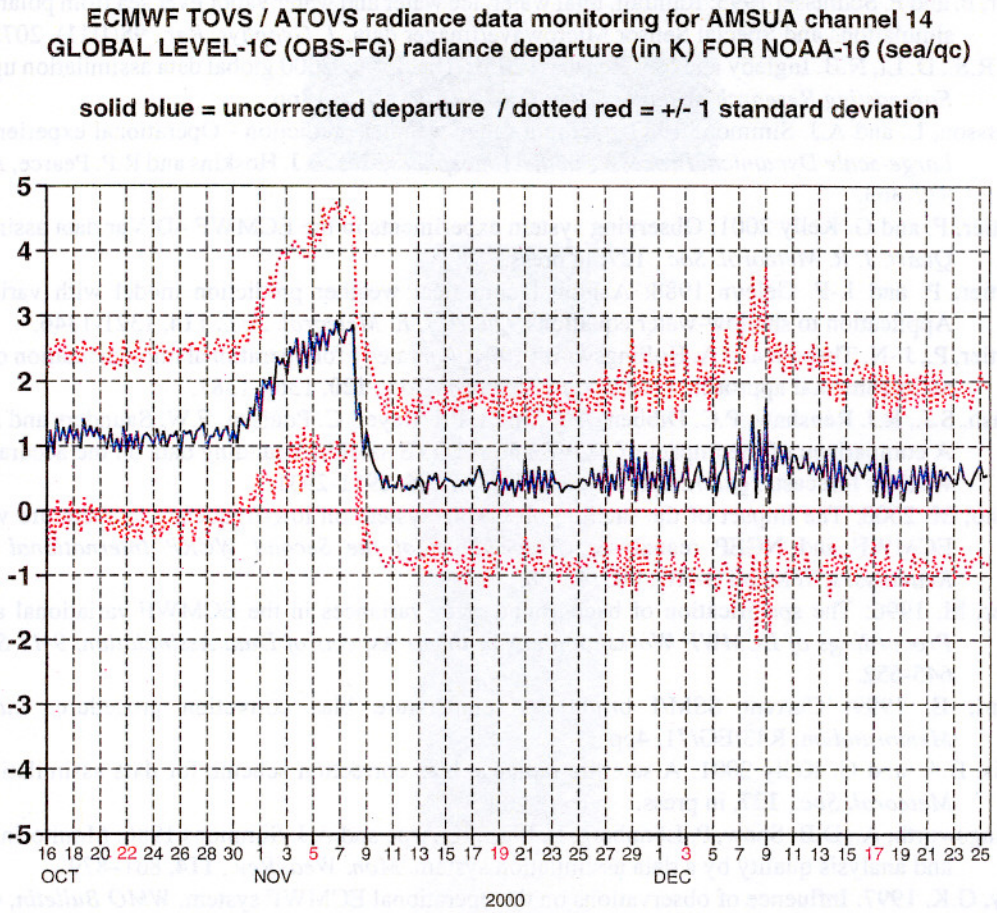

Fig. 16 Differences between observed and simulated background values of brightness temperature (K) for channel 14 of the AMSU-A instrument of NOAA-16, averaged over all observations processed for each assimilation cycle from 16 October to 26 December 2000. The solid blue line shows mean differences, and the dotted red lines show the mean plus and minus one standard deviation of the differences.

There is good reason to be optimistic that further significant improvements in forecasts will arise from the assimilation of satellite data. Better exploitation of existing types of data should come in particular from a more comprehensive use of data over land and ice-covered surfaces and from utilization of the information the data contain on clouds and precipitation. The forthcoming generation of high-resolution infrared sounders, beginning with AIRS and IASI, offer exciting prospects for their use. Direct assimilation of radiances from improved instruments in geostationary orbit such as SEVIRI on Meteosat Second Generation and from the improved SSMIS microwave instrument on the DMSP satellites are likely to bring further benefits, quite apart from the benefits expected from a variety of other new instruments several of which are discussed elsewhere in these Proceedings.

Acknowledgements

Erik Andersson, Lars Isaksen, Per Kållberg, Graeme Kelly, Tony McNally, Jean-Francois Mahfouf and Jean-Noël Thépaut are acknowledged with thanks for providing information and results for presentation in this paper.

References

Alishouse, J.C., S.A. Snyder, J. Vongsathorn and R.R. Ferraro, 1990: Determination of oceanic total precipitable water from the SSM/I. *IEEE Trans. Geosci. Remote Sensing*, **28**, n5, 811-816.

Andersson, E., M. Fisher, R. Munro and A. McNally 2000: Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Quart. J. R. Meteorol. Soc.*, **126**, 1455-1472.

Andersson, E. and H. Järvinen 1999: Variational quality control. *Quart. J. R. Meteorol. Soc.*, **125**, 697-722.

Bauer, P. and P. Schlüssel 1993: Rainfall, total water, ice water and water vapor over sea from polarized microwave simulations and Special Sensor Microwave/Imager data. *J. Geophys. Res.*, **98(D11)**, 20737-20759.

Bell, R.S., D. Li, N.B. Ingleby and R.J. Renshaw 2001: The Spring 2000 global data assimilation upgrade package. *Forecasting Research Technical Report, Met Office*, **327**, 18pp.

Bengtsson, L. and A.J. Simmons 1983: Medium-range weather prediction - Operational experience at ECMWF. *Large-scale Dynamical Processes in the Atmosphere*, Eds. B.J. Hoskins and R.P. Pearce, Academic Press, 337-363.

Bouttier, F. and G. Kelly 2001: Observing system experiments in the ECMWF 4D-Var data assimilation system. *Quart. J. R. Meteorol. Soc.*, **127**, in press.

Courtier, P. and J.-F. Geleyn 1988: A global numerical weather prediction model with variable resolution: Application to shallow-water equations. *Quart. J. R. Meteorol. Soc.*, **114**, 1321-1346.

Courtier, P., J.-N. Thépaut and A. Hollingsworth 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. R. Meteorol. Soc.*, **120**, 1367-1387.

English, S.J., R.J. Renshaw, P.C. Dibben, A.J. Smith, P.J. Rayer, C. Poulsen, F.W. Saunders and J.R. Eyre, 2000: A comparison of the impact of TOVS and ATOVS satellite sounding data on the accuracy of numerical weather forecasts. *Quart. J. R. Meteorol. Soc.*, **126**, 2911-2931.

Fiorino, M. 2000: The impact of the satellite observing system on low-frequency temperature variability in the ECMWF and NCEP reanalyses. *Proceedings of the Second WCRP International Conference on Reanalyses*, WMO/TD-NO. 985, 65-68.

Fisher, M. 1996: The specification of background error variances in the ECMWF variational analysis system. *Proceedings of ECMWF Workshop on Non-linear Aspects of Data Assimilation, 9-11 September 1996*, 645-652.

Gérard, E. 1999: Current SSM/I brightness temperature bias correction procedure. *Internal ECMWF Memorandum*, R43/EG/71, 4pp.

Harris, B.A. and G. Kelly 2001: A satellite radiance bias correction scheme for data assimilation. *Quart. J. R. Meteorol. Soc.*, **127**, in press.

Hollingsworth, A., D.B. Shaw, P. Lönnberg, L. Illari, K. Arpe and A.J. Simmons 1986: Monitoring of observation and analysis quality by a data assimilation system. *Mon. Wea. Rev.*, **114**, 861–879.

Kelly, G.K. 1997: Influence of observations on the operational ECMWF system. *WMO Bulletin*, **46**, 336-342.

Lorenc, A.C., S.P. Ballard, R.S. Bell, N.B. Ingleby, P.L.F. Andrews, D.M. Barker, J.R. Bray, A.M. Clayton, T. Dalby, D. Li, T.J. Payne and F.W. Saunders 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. R. Meteorol. Soc.*, **126**, 2991-3012.

Mahfouf, J.-F. and F. Rabier 2000: The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. *Quart. J. R. Meteorol. Soc.*, **126**, 1171-1190.

McNally, A.P., E. Andersson, G. Kelly and R.W. Saunders 1999: The use of raw TOVS/ATOVS radiances in the ECMWF 4D-Var assimilation system. ECMWF Newsletter, 83, 2-7.

McNally, A.P., J.C. Derber, W. Wu and B.B. Katz 2000a: The use of TOVS level-1b radiances in the NCEP SSI analysis system. *Quart. J. R. Meteorol. Soc.*, **126**, 689-724.

McNally, A.P., G. Kelly and A. Untch 2000b: Assimilation of AMSU channel 14 radiances. *Internal ECMWF Memorandum*, R43.8/AM/55, 20pp.

Marécal, V. and J.-F. Mahfouf, 2000: Revised quality control for SSM/I 1D-Var. *Internal ECMWF Memorandum*, R46.2/VM/1, 14pp.

Okamoto, K. and H. Tada 2001: Recent developments in assimilation of TOVS and ATOVS at JMA. To appear in *Proceedings of 11th International ATOVS Study Conference*.

Simmons, A.J., R. Mureau and T. Petroliagis 1995: Error growth and predictability estimates for the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739-1771.

Stoffelen, A. and D. Anderson 1997: Ambiguity removal and assimilation of scatterometer data. *Quart. J. Roy. Meteor. Soc.*, **123**, 491-518.

Talagrand, O. and P. Courtier 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. R. Meteorol. Soc.*, **113**, 1311-1328.