IFS DOCUMENTATION – Cy31r1
Operational implementation 12 September 2006

PART II: DATA ASSIMILATION

Table of contents

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Chapter 1
Overview

Table of contents

1.1 INTRODUCTION

This documentation of 4D-Var and 3D-FGAT serves as a scientific guide to the data assimilation codes within ECMWF's Integrated Forecasting System (IFS). The documentation is divided into a number of chapters. This first chapter provides an overview of ECMWF's data assimilation system, a brief history of its evolution and a description of the basic scientific formulation.

The second chapter describes the practical implementation of the multi-resolution incremental method for 4D-Var data assimilation (used operationally since January 2003, Cy25r3), and the solution algorithm including minimization and preconditioning. The 3D-FGAT configuration as used in the production of ERA-40 (Uppala *et al.*, 2005) and operationally in the so-called Boundary-Condition suite is described in Chapter 3. Thereafter follows a description of the background term (Chapter 4) and a chapter on observation operators and the organisation of observation cost-function calculations for conventional and satellite data (Chapter 5). Chapter 6 deals with the computation and cycling of background and analysis errors and Chapter 7 is on initialization and control of gravity waves. Diagnostic tools for investigation of the performance of the assimilation system are described in Chapter 8. The modules for observation selection, quality control and screening are described in Chapter 9 and Chapter 10. Chapter 11 outlines the snow analysis, Chapter 12 describes the soil analysis, Chapter 13 describes the sea surface temperature and sea-ice analysis. Finally Chapter 14 provides summary information about the main unix-scripts and important files, and a schematic of the data flow between the various jobs steps that constitute a data assimilation cycle.

1.2 SCIENTIFIC PUBLICATIONS

The scientific description of 3D/4D-Var has been published in a series of papers in the *Quarterly Journal of the Royal Meteorological Society* (QJRMS), in ECMWF workshop proceedings and Technical Memoranda over the years. The incremental formulation was introduced by Courtier *et al.* (1994). The ECMWF implementation of 3D-Var was published in a three-part paper by Courtier *et al.* (1998), Rabier *et al.* (1998) and Andersson *et al.* (1998). The original observation operators for conventional data can be found in Vasiljevic *et al.* (1992). The methods for assimilation of TOVS radiance data and ERS scatterometer data were developed by Andersson *et al.* (1994) and Stoffelen and Anderson (1997), respectively. The pre-operational experimentation with 4D-Var has been documented in three papers by Rabier *et al.* (2000), Mahfouf and Rabier (2000) and Klinker *et al.* (2000). The background term has been published by Derber and Bouttier (1999), with more recent developments described by Fisher (2003).

Papers of data usage and impact include a study of commercial aircraft data (Cardinali *et al.*, 2003), scatterometer impact (Isaksen, 1997; Isaksen and Stoffelen, 2000; Isaksen and Janssen, 2004), conventional and satellite humidity data (Andersson *et al.*, 2004), ozone analysis (Dethof and Holm, 2004), time series of frequent data (Andersson *et al.*, 2001), wind profiler data (Bouttier, 2001b; Andersson and Garcia-Mendez, 2002), TOVS radiance data (McNally *et al.*, 1999), water-vapour

radiances from geostationary satellites (Munro *et al.*, 2004; Köpken *et al.*, 2004), and the use of high-volume satellite data in general (Thépaut and Andersson, 2003).

1.3 BRIEF HISTORY OF 3D- AND 4D-VAR IN ECMWF OPERATIONS

3D-Var was implemented in ECMWF operations on 30 January 1996, and was replaced on 25 November 1997 by 4D-Var. The two three-part series of QJRMS papers mentioned above basically describe the two schemes as they were at the time of their implementation. However, there have been very significant developments of the variational data assimilation system during its time in operations. Major code-upgrades have taken place in connection with migrations to new computers. 3D-Var was first implemented on a CRAY C90 shared-memory parallel computer system and was migrated to a distributed-memory Fujitsu VPP700 machine in September 1996. Further migrations of the codes took place in May 2000 to a Fujitsu VPP5000, and in March 2003 to an IBM P690, and later in 2004 to an IBM P690+. A history of the evolution of ECMWF super-computing is available at www.ecmwf.int. The current IBM implementation of the assimilation system utilizes both shared-memory (OpenMP) and distributed-memory (MPI) parallelism. A general list of changes to the operational forecasting system is available at http://www.ecmwf.int/products/data/technical/model_id/index.html.

The observation handling and data screening modules have also developed rapidly to keep pace with the changing computer environment, and the arrival of new observational data sets. The codes inherited from OI were replaced with new codes at the time of the migration to Fujitsu in 1996. The need for improved parallelism and rapidly increasing data volumes lead to the development of the Observation DataBase (ODB) software (see ODB documentation), facilitating in March 2003 the move to the IBM computers and the introduction of high-spectral resolution satellite data (e.g. AIRS). The quality control, data selection and screening algorithms are described in the paper by Järvinen and Undén (1997), and variational quality control of observations in Andersson and Järvinen (1999).

One of the most important aspects of a variational assimilation is the specification of background errors. The original formulation of Courtier *et al.* (1998) was replaced in May 1997 by that of Derber and Bouttier (1999). The latter formulation is still used as described in Chapter 4, including more recent improvements such as non-linear balance (January 2003, Cy25r3, Fisher (2003)), wavelet-Jb (April 2005, Cy29r1, Fisher (2006)), introduction of ozone as an analysis variable (Dethof and Holm, 2004) in October 1999 (Cy21r4), the new humidity analysis (Hólm *et al.*, 2002) in October 2003 (Cy26r3), and Jb for GEMS variables (Benedetti and Fisher, 2006). The cycling algorithms for analysis and background errors (Fisher and Courtier, 1995, and Chapter 6) were introduced in 1996. The calibration of background error statistics is since October 1999 based on an ensemble of 3D-Var data assimilations, updated in January 2003 to statistics based on a 4D-Var ensemble (Fisher, 2003).

On the 25th November 1997 6-hourly 4D-Var was introduced operationally, at resolution T213L31, with two iterations of the outer loop: the first at T63L31 with 50 iterations (simplified physics) and the second with 20 iterations (with tangent-linear physics, at same resolution). In April 1998 the resolution was changed to T_L319 and in June 1998 we revised the radiosonde/pilot usage (significant levels, temperature instead of geopotential) and we started using time-sequences of data (Järvinen *et al.*, 1999), so-called 4D-screening. The data assimilation scheme was extended higher into the atmosphere on 10 March 1999, when the $T_L319L50$ model was introduced, which in turn enabled the introduction in May 1999 of ATOVS radiance data (McNally *et al.*, 1999). In October 1999 the vertical resolution of the boundary layer was enhanced taking the number of model levels to a total of L60. In summer 2000 the 4D-Var period was extended from 6 to 12 hours (Bouttier, 2001a), whereas the ERA-40 configuration (Uppala *et al.*, 2005) was built as an FGAT (first guess at the appropriate time) of 3D-Var with a period of 6 hours (Chapter 3). In November 2000, the horizontal resolution of 4D-Var was increased to $T_L511L60$, with inner loop resolution enhanced from T63L60 to $T_L159L60$ using the linearized semi-Lagrangian scheme. In February 2006, the horizontal resolution of 4D-Var was increased to $T_L799L90$, with inner loop resolution enhanced to $T_L255L90$

In January 2003 the 4D-Var solution algorithm was comprehensively revised (in Cy25r4, Andersson *et al.*, 2003) to take advantage of the efficiency of conjugate gradient minimisation (with pre-conditioning, as introduced one year earlier) and a multi-resolution incremental technique (Trémolet, 2005). This is the

4D-Var solution algorithm that is presented here in Chapter 2. The early delivery suite was introudiced in June 2004 (in Cy28r2, Haseler, 2004).

1.4 INCREMENTAL FORMULATION OF VARIATIONAL DATA ASSIMILATION

In 3D/4D-Var an objective function J is minimized. The cost function consists of three terms:

$$J = J_{\mathrm{b}} + J_{\mathrm{o}} + J_{\mathrm{c}} \tag{1.1}$$

measuring, respectively, the discrepancy with the background (a short-range forecast started from the previous analysis), J_{b}, with the observations, J_{o} and with the slow character of the atmosphere, J_{c}. The J_{c}-term controls the amplitude of fast waves in the analysis and is described in Chapter 7. It is omitted from the subsequent derivations in this section.

In its incremental formulation (Courtier *et al.*, 1994), we write

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}) \tag{1.2}$$

$\delta\mathbf{x}$ is the increment and at the minimum the resulting analysis increment $\delta\mathbf{x}^{\mathrm{a}}$ is added to the background $\mathbf{x}^{\mathrm{b}}$ in order to provide the analysis $\mathbf{x}^{\mathrm{a}}$ given by

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \delta\mathbf{x}^{\mathrm{a}} \tag{1.3}$$

$\mathbf{B}$ is the covariance matrix of background error while $\mathbf{d}$ is the innovation vector

$$\mathbf{d} = \mathbf{y}^{\mathrm{o}} - H\mathbf{x}^{\mathrm{b}} \tag{1.4}$$

where $\mathbf{y}^{\mathrm{o}}$ is the observation vector. $\mathbf{H}$ is a suitable low-resolution linear approximation of the observation operator H in the vicinity of $\mathbf{x}^{\mathrm{b}}$, and $\mathbf{R}$ is the covariance matrix of observation errors. The incremental formulation of 3D/4D-Var consists therefore of solving for $\delta\mathbf{x}$ the inverse problem defined by the (direct) observation operator $\mathbf{H}$, given the innovation vector $\mathbf{d}$ and the background constraint. The gradient of J is obtained by differentiating (1.2) with respect to $\delta\mathbf{x}$ to give

$$\nabla J = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})\delta\mathbf{x} - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d} \tag{1.5}$$

At the minimum, the gradient of the objective function vanishes, thus from (1.5) we obtain the classical result that minimizing the objective function defined by (1.2) is a way of computing the equivalent matrix-vector products given by

$$\delta\mathbf{x}^{\mathrm{a}} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{d} \tag{1.6}$$

where $\mathbf{B}$ and $\mathbf{R}$ are positive definite, see e.g. Lorenc (1986) for this standard result. $\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ may be interpreted as the square matrix of the covariances of background errors in observation space while $\mathbf{B}\mathbf{H}^{\mathrm{T}}$ is the rectangular matrix of the covariances between the background errors in model space and the background errors in observation space. 3D-Var uses the observation operator $\mathbf{H}$ explicitly and statistical models ($\mathbf{B}$ and $\mathbf{R}$) are required only for describing the statistics of the background errors in model space and the observation error in observation space. Consequently, in 3D/4D-Var it is easy, from an algorithmic point of view, to make use of observations such as TOVS radiances, which have quite a complex dependence on the basic analysis variables. The background term and background error covariance modelling (i.e. $\mathbf{B}$) are described in Chapter 4 and observation operators $\mathbf{H}$ are described in Chapter 5.

Chapter 2
4D variational assimilation

Table of contents

2.1 INTRODUCTION

This part of the documentation covers the top level controls of 4D-Var. It gives a detailed description of the solution algorithm, and the various steps that are performed during a 4D-Var simulation (SIM4D). The procedure consists of nested iterations, called inner and outer loops, which communicate through files that either contain meteorological fields, observations, or parameters. The data flow is documented in Chapter 14. All of this chapter also applies to 3D-Var with some exceptions which will be detailed in Chapter 3. The interpolation of model fields to observation points (OBSHOR) and the organization of the data in memory (yomsp, yommvo) is described in Chapter 5. The structure of the computation of the observation cost function and its gradient, managed by the routines OBSV and TASKOB can also be found in Chapter 5. The background term and the 'change-of-variable' operators are explained in Chapter 4.

2.2 SOLUTION ALGORITHM

2.2.1 The incremental method

The adopted solution algorithm is incremental (Courtier *et al.*, 1994), which allows for considerable flexibility with respect to the computer expense of the various job-steps of the 4D-Var minimisation. In the

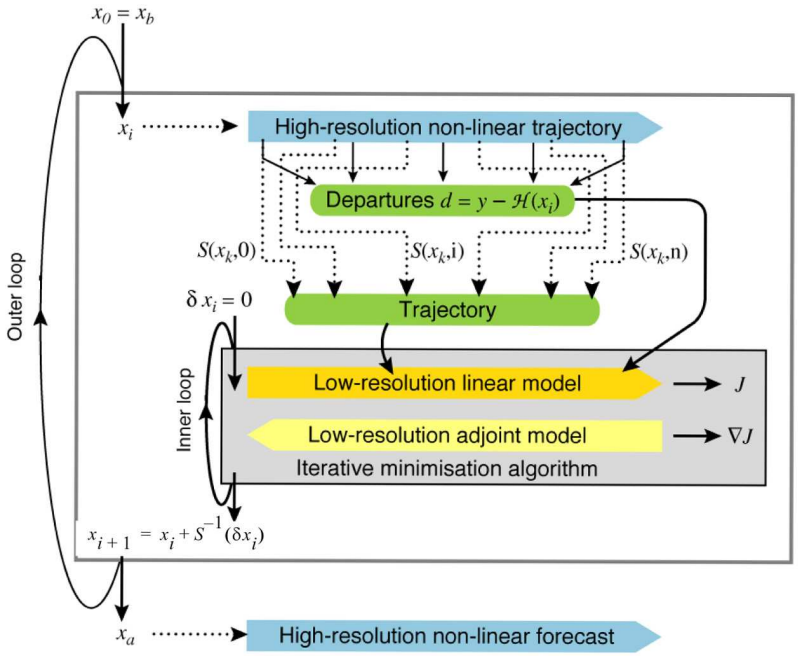

Figure 2.1 *Schematic of the revised 4D-Var solution algorithm implemented in January 2003 (Cy25r4). Outer loops are performed at high resolution (currently T799) using the full non-linear model. Inner iterations are performed at lower resolution (first T95, then T255) using the tangent-linear forecast model, linearised around a 12-hour succession of model states ('the trajectory') obtained through interpolation from high resolution (S denotes the truncation operator, J the cost function and x the atmospheric state vector).*

incremental approach the highest possible resolution is used for the computation of the model trajectory, and for calculating the departures between observations and model, whereas a lower-resolution model (its adjoint and tangent linear) are used for the iterative and relatively costly computation of analysis increments (Trémolet, 2004; Radnóti *et al.*, 2005). The lower-resolution iterations (the inner-loops) can optionally be nested within a set of outer-loop iterations at full resolution (Trémolet, 2005). Apart from the resolution, the cost of the inner-loops will depend also upon the complexity of the inner-loop model, e.g. the use of simpler or more complete representations of the physical processes (Janisková *et al.*, 2002; Tompkins and Janisková, 2004).

In a further 'multi-resolution' extension to the incremental method (Veerse and Thépaut, 1998) the inner-loop resolution is increased with each iteration of the outer-loop. A schematic is shown in Fig. 2.1. In particular, the information about the shape of the cost-function obtained during the early low-resolution iterations provides a very effective pre-conditioner (Chapter 6) for subsequent iterations at higher-resolution, thus reducing the number of costly iterations. The inner-loops can be particularly efficiently minimised using the conjugate gradient method, provided the cost-function is purely quadratic (Fisher, 1998), i.e. the operators involved in the definition of the cost function (the model and the observation operators) are purely linear. For this reason, the inner-loops have been made entirely linear, with the non-linear effects gathered at the outer-loop level. The convergence properties of the outer-loop iterations have been investigated by Trémolet (2005).

2.2.2 The job-steps

In the Cy31r1 operational configurations the assimilation window is 12-hours long, running from 09–21 UTC to produce the 12 UTC analysis and forecast products, and from 21–0R9UTC for the 00 UTC production (Haseler, 2004). Several different job steps are performed.

(i) Comparison of the observations with the background at high resolution to compute the innovation vectors. These are stored in the NCMIFC1-word of the ODB (the observation database) for later

use in the minimization. This job step also performs **screening** (i.e. blacklisting, thinning and quality control against the background) of observations (see Chapter 10). The screening determines which observations will be passed for use in the main minimisation. Very large volumes of data are present during the screening run only, for the purpose of data monitoring. The model trajectory is interpolated to the resolution of the next job step and written out.

(ii) First minimization at low resolution to produce preliminary low-resolution analysis increments, using simplified tangent-linear physics, and tangent-linear observation operators. The eigenvectors of the analysis Hessian are computed and these will be used to precondition subsequent inner-loop iterations. The analysis and background errors required for the cycling of error variances are calculated, as described in Chapter 6.

(iii) Update of the high-resolution trajectory to take non-linear effects partly into account. Observed departures from this new atmospheric state are stored in the ODB and the analysis problem is re-linearized around the updated model state. Variational quality control is computed, and the resulting QC weights will be used in the subsequent minimisation. The updated model trajectory is interpolated to the resolution of the next job step and written out.

(iv) Second main minimization at increased resolution with more complete representation of tangent-linear physics.

(v) Formation of the high-resolution analysis (described below) and a comparison of the analysis with all observations (including those not used by the analysis, for diagnostic purposes).

Each of the job steps is carried out by a different configuration of IFS. They are commonly called as follows.

(i) **The first trajectory run:** includes a model integration, comparison to observations, and observation screening (quality control and data selection) and is sometimes called **the screening run.** NCONF = 2, LSCREEN = .TRUE.

(ii) **The first minimization:** uses simplified physics, typically run at low resolution. This job step includes estimation of analysis and forecast error variances (required for cycling), and calculation of Hessian eigenvectors for pre-conditioning of subsequent minimisation(s). NCONF = 131, LSPHLC = .TRUE., LAVCGL = .TRUE.

(iii) **The first trajectory update:** applies the analysis increments obtained in the first minimisations and performs another forecast integration with comparison to observations. This provides a new linearisation state for the next minimisation. NCONF = 1, LOBS = .TRUE.

(iv) **The second minimization:** uses more complete tangent-linear physics, typically higher resolution increments. NCONF = 131, LSPHLC = .FALSE.

(v) **Late 4D-start runs:** if LATE4DSTART (default), the analyses (type = an) at the main synoptic hours (00, 06, 12, 18), that fall within the assimilation window, are formed in separate quick job-steps by adding the low-resolution increment to the penultimate high-resolution trajectory of the corresponding time (no forecast integration). See also schematic in Chapter 14.

(vi) **The final trajectory runs:** carries out verification screening – that is comparison between observations and final analysis. In the final trajectory job-steps with NUPTRA = 999 the final analysis (type = 4v) is formed (Bouttier, 2001a), by adding the low-resolution increment to the background (at initial time), and integrating to the analysis times. NCONF = 1, LOBS = .TRUE., NUPTRA = NRESUPD

The steps (iii) and (iv) are referred to as the second iteration of the *outer loop*, and these can optionally be iterated further to incorporate additional non-linear effects. The trajectory update is not normally done in 3D-Var. The *inner loop* takes place within the main minimization, job steps (ii) and (iv).

2.2.3 Interpolation of trajectory and increments

A truncation operator (shown as S in the schematic, Fig. 2.1 above) is required to take the trajectory fields from high to low resolution. This is done using appropriate grid-point interpolations for the surface grid-point fields, whereas upper-air fields are truncated spectrally. Humidity and ozone (and any other grid-point fields) are interpolated in grid-point space (either linear or cubic) to the resolution of the inner loops. Initial time model cloud fields are interpolated to lower resolution using the so-called 'full-pos'

configuration of IFS. Trajectory cloud fields and trajectory of physical tendencies are generated in a (low resolution) model integration in which the model state is replaced by the interpolated trajectory at each time instance for which it is available. The trajectory handling is managed by the module TRAJECTORY.

The spectral and grid-point analysis increments produced by the minimisation are read in by the routine RDFPINC. The spectral fields are padded with zeroes for the wave numbers not represented in the inner-loops. The increments for the spectral model variables (vorticity, divergence and logarithm of surface pressure) are added to the background fields. The temperature increments are added to the temperature of the background (after it has been converted from virtual temperature) and the resulting temperature analysis is converted back to virtual temperature. The humidity and ozone increments are read in as grid-point fields and interpolated to the outer-loop resolution and added to the background. Checks for negative values are applied to humidity, ozone (and trace gasses).

For the 91-level model extension into the mesosphere, it was found that the TL model was unstable in certain situation. This was resolved by introducing smooting of the trajectory fields, if the inner-loop resolution is T255, or higher. This is implemented as a fourth-order diffusion filter applied to the spectral trajectory fields, reducing the amplitude of the shortes wave by a factor five. This is done in TRAJ_MAIN just before the fields are written to files in ifstraj. The filtering is controlled by the variables FILTERFACTOR, FILTEREXPO, FILTERRESOL in NAMVAR, available also in prepIFS. In addition, it was found necessary to reduce the amplitude of the analysis increments near the top of the L91 model. They are tapered to zero applying the factor REDINC, over the topmost NLEV_REDINC=4 levels (NAMJG), in RDFPINC.

2.2.4 Pre-conditioning and control variable

In practice, it is necessary to pre-condition the minimization problem in order to obtain a quick convergence. As the Hessian (the second derivative) of the objective function is not accessible, Lorenc (1988) suggested the use of the Hessian of the background term J_b. The Hessian of J_b is the matrix $\mathbf{B}$. Such a preconditioning may be implemented either by a change of metric (i.e. a modified inner product) in the space of the control variable, or by a change of control variable. As the minimization algorithms have generally to evaluate several inner products, it was found more efficient to implement a change of variable (under CHAVAR, CHAVARIN, etc.). Algebraically, this requires the introduction of a variable χ such that

$$J_b = \chi^T \chi \tag{2.1}$$

Comparing (1.2) and (2.1) shows that $\chi = \mathbf{B}^{-1/2}\delta\mathbf{x}$ satisfies the requirement. χ thus becomes the *control variable* of the preconditioned problem. This is indeed what has been implemented, as will be explained in Section 4.2. A single-observation analysis with such preconditioning converges in one iteration.

The J_b-based pre-conditioning is sufficient when the B-term dominates over the observation term of the Hessian. With increasing amounts of observational information, and in cases with locally dense observation coverage, the observation term can be dominant over the background term in defining the shape of the cost-function (its second derivative). The combined Lanczos/conjugate gradient method allows computation of the leading eigenvectors and eigenvalues of the Hessian while solving for the analysis, essentially at no extra cost. This method is fully described in Chapter 6, and it is used for all inner-loop iterations. The Hessian eigenvector information obtained at low resolution is used as pre-conditioner at subsequent inner-loop iterations at higher resolution. This has been shown to be a very effective way of reducing the number of iterations required at higher inner-loop resolutions.

2.2.5 Minimization

The minimization problem involved in 3D/4D-Var can be considered as large-scale, since the number of degrees of freedom in the control variable is of the order of 10^7. In the case of purely quadratic cost-function (as in operational 4D-Var) the most efficient minimisation is provided by the conjugate gradient algorithm (CONGRAD called from CVA2 and FORECAST_ERROR). Congrad is preconditioned with the leading Hessian eigenvectors, as described in Chapter 6. The forecast error configuration (Chapter 6, switch LAVCGL = .TRUE. in yomvar) always uses CONGRAD, and otherwise it is activated by the switch LCONGRAD in yomvar.

For non-quadratic problems, an efficient descent algorithm was provided by the Institut de Recherche en Informatique et Automatique (INRIA, France). It is a variable-storage quasi-Newton algorithm (M1QN3) described in Gilbert and Lemaréchal (1989). M1QN3 uses the available in-core memory to update an approximation of the Hessian of the cost function. In practice, ten updates (NMUPD, namiomi) of this Hessian matrix were used. M1QN3 is currently used in the 3D-FGAT configuration, described in Chapter 3.

2.3 TOP-LEVEL CONTROLS

The routines CVA1, CVA2 and FORECAST_ERROR control the variational configurations of IFS. The flow diagram of CVA1 and CVA2 is shown in Fig. 2.2. The sepctral and grid-point first guess fields (FG) have been read in to the SP7-arrays and GP7-arrays (in YOMSP) by SUECGES, called from SUJBSTD within the J_b setup, see Subsection 4.3.3. At the start of CVA1 additional setups for the variational configurations are done (SU1YOM). The SP3-arrays and GFL-arrays, i.e. the current model state, (in YOMSP) are filled by a call to the setup routine SUINIF in the routine SUVAZX, and only for the first minimisation job-step. For subsequent minimisations the initial state for analysed variables must come from the previous minimisation, i.e. a warm start. This is achieved by getting the control vector YVAZX by calling GETMINI and performing an inverse change of variables (calling CHAVARIN) to transform the control vector to model variables (SP3/SP2-arrays and GFL-arrays). Non-analysed fields (e.g. cloud fields) are read by a call to SUINIF earlier in SUVAZX.

2.3.1 Last simulation

After minimisation is complete one final simulation is performed in CVA2 or FORECAST_ERROR. This simulation is diagnostic, and characterized by the simulation counter being set to 999, NSIM4D = NSIM4DL, yomvar. The observation departure from the low-resolution analysis, $\mathbf{y}^o - H\mathbf{x}^a_{LR}$, is computed and stored in the NCMIOMN-word of the ODB. Finally at the end of CVA2 and FORECAST_ERROR, respectively, the updated ODB is written to disk, using the routine WRITEOBA.

2.4 A SIMULATION

2.4.1 The cost function and its gradient

A simulation consists of the computation of J and ∇J. This is the task of the routine SIM4D, see Fig. 2.3 for the flow diagram. The input is the latest value of the control variable χ in the array VAZX, computed by M1QN3, or CONGRAD. First J_b and its gradient are computed (see Section 4.2) using

$$\begin{aligned} J_b &= \chi^T\chi \\ \nabla_\chi J_b &= 2\chi \end{aligned} \tag{2.2}$$

The gradient of J_b with respect to the control variable is stored in the array VAZG (YOMCVA).

(i) Copy χ from VAZX to SP3-arrays (YOMSP) using the routine CAIN.
(ii) Compute $\mathbf{x}$, the physical model variables, using CHAVARIN so that

$$\mathbf{x} = \delta\mathbf{x} + \mathbf{x}^b = L\chi + \mathbf{x}^b \tag{2.3}$$

(iii) If L131TL then work in terms of increments. SUBFGS does $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^g$, where $\mathbf{x}^g$ is the guess (not necessarily equal to the background).
(iv) Perform the direct integration of the model (if 4D-Var), using the routine CNT3 or CNT3TL, and compare with observations. See Section 2.5.
(v) Calculate J_o for which OBSV is the master routine.
(vi) Perform the adjoint model integration (if 4D-Var) using CNT3AD, and observation operators' adjoint.
(vii) Calculate $\nabla_\mathbf{x} J_o$, and store it in SP3.
(viii) J_c and its gradient are calculated in COSJC called from CNT3AD, if LJC is switched on (default) in namvar.

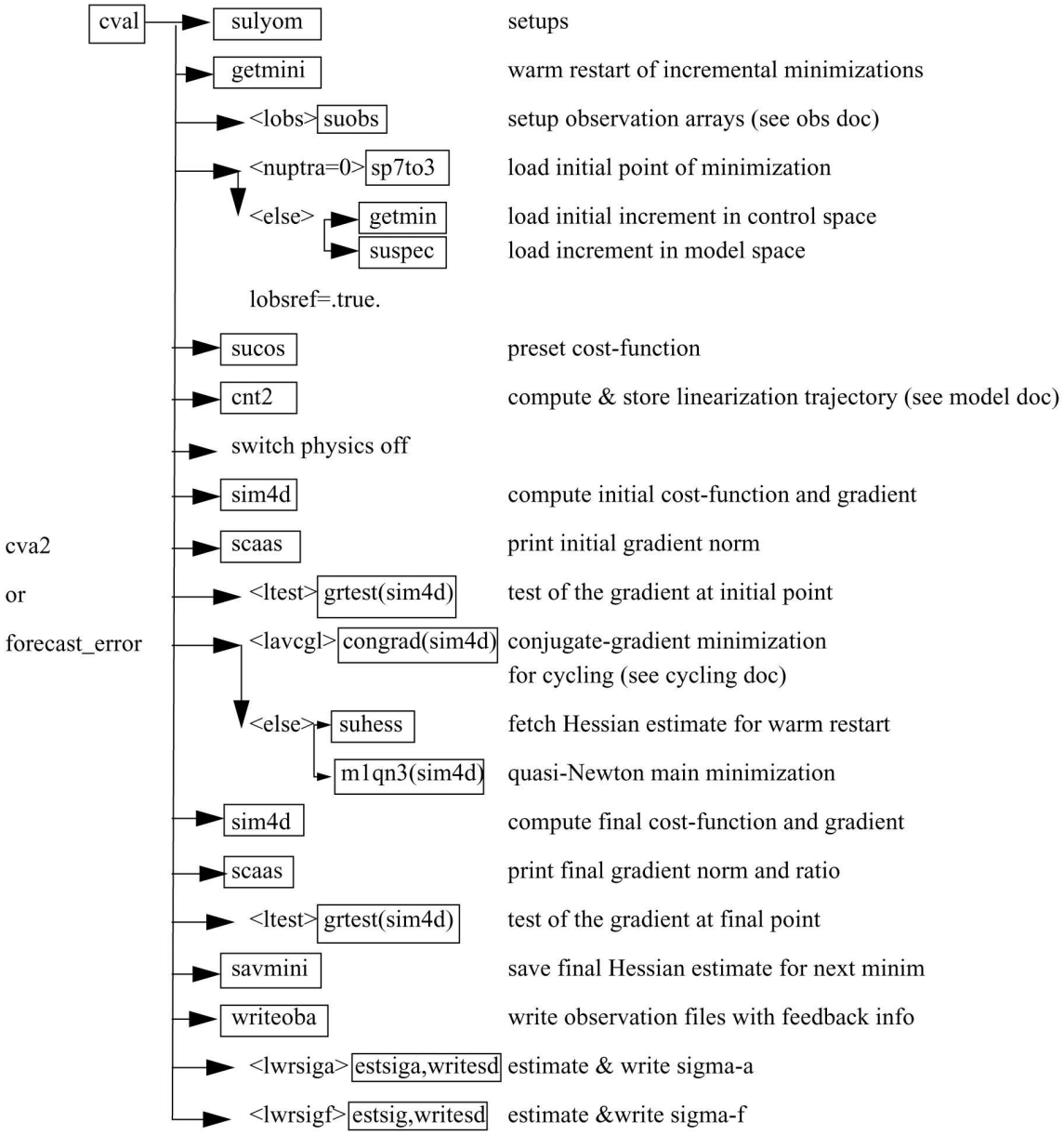

Figure 2.2 *Flow diagram for subroutines CVA1 and CVA2.*

(ix) Transform $\nabla_{\mathbf{x}} J_o + \nabla_{\mathbf{x}} J_c$ to control variable space by applying CHAVARINAD.
(x) Copy $\nabla_{\chi} J_o + \nabla_{\chi} J_c$ from SP3 and add to $\nabla_{\chi} J_b$, already in the array VAZG, using CAIN.
(xi) Add the various contributions to the cost function together, in EVCOST, and print to log file using prtjo.
(xii) Increase the simulation counter NSIM4D by one.

The new J and $\nabla_{\chi} J$ are passed to the minimization algorithm to calculate the χ of the next iteration, and so on until convergence (or the maximum number of iterations) has been reached.

2.4.2 Interface between control variable and model arrays

The purpose of the routine CAIN (the canonical injection) is to identify those parts of the model state that should be included in the variational control variable. This is controlled by on/off switches such as NVA2D and NVA3D (yomcva) initialized in SUALCTV. The scalar product used is the one defined by

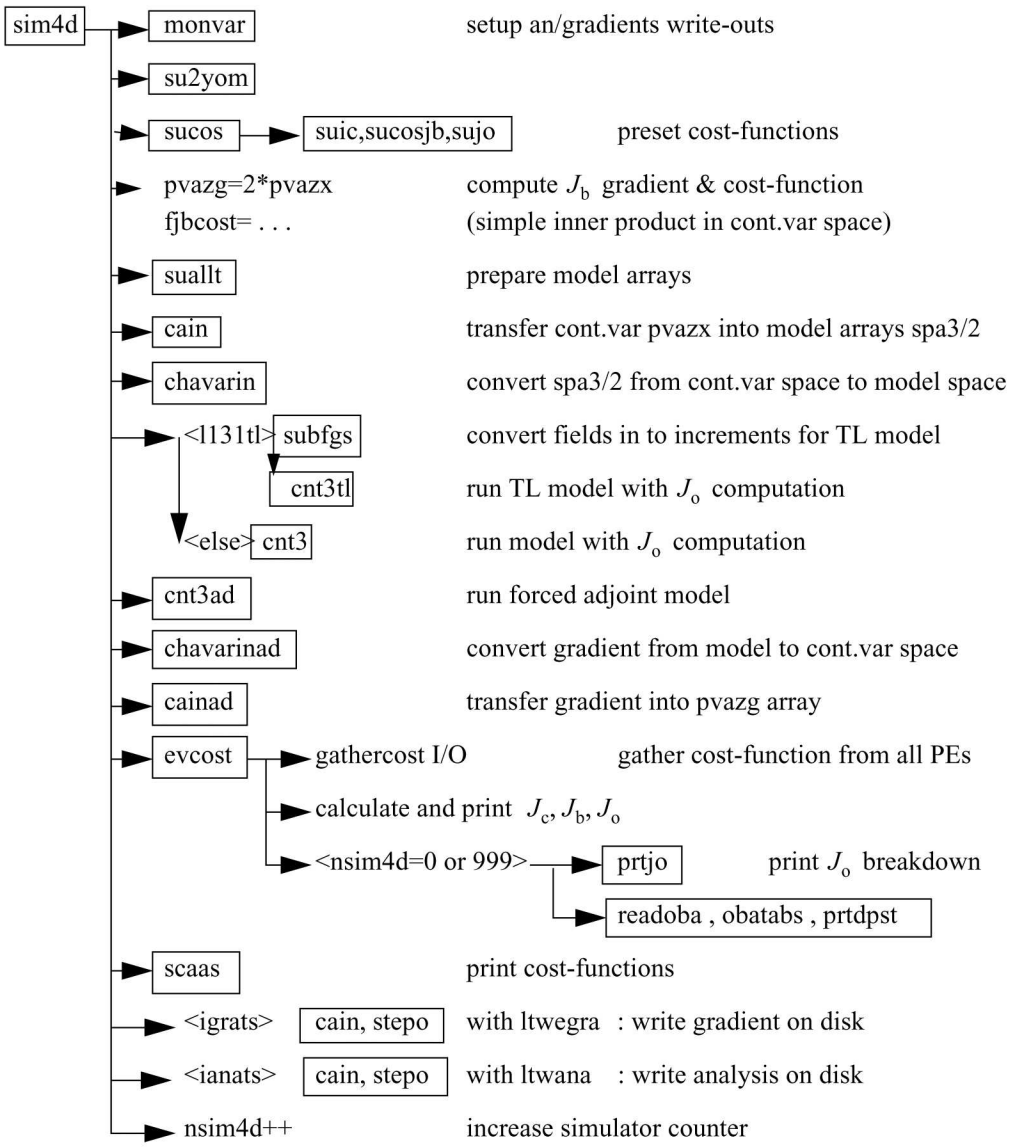

Figure 2.3 *Flow diagram for the subroutine sim4d.*

the array SCALP (in yomcva, set up in the routine SCALJGS called from SUSCAL), which is 1 if $m = 0$, and 2 otherwise. This allows the compression of arrays of the type VAZX while using the L^2 norm on the sphere with real fields in spectral space.

The control vector currently consists of four parts, as defined in the module CONTROL_VECTORS residing in the ifsaux library. Any of the following four parts can be absent.

(i) The initial-condition control-variables that define the original 4D-Var problem.
(ii) A representation of model error (for weak-constraint 4D-Var), see Trémolet (2003).
(iii) The TOVS control vector (LTOVSCV), used to estimate surface skin temperature at the radiance field-of-view locations.
(iv) Components for the limited-area model ALADIN.

The organisation of elements of the control-vector is now quite flexible and allows for future planned extensions, such as observation bias coefficients (e.g. for radiance data, and wavelet-J_b components).

2.5 THE ASSIMILATION WINDOW

4D-Var is a temporal extension of 3D-Var. Observations are organized in time-slots (currently half-hourly) as described in the ODB documentation. The cost-function measures the distance between a model trajectory and the available information (background, observations) over an assimilation interval or window. For a 12-hour window (in the configuration used before June 2004), it is either (03 UTC–15 UTC) or (15 UTC–03 UTC). Since then it is either (09 UTC–21 UTC) or (21 UTC–09 UTC), see Haseler (2004). Equation (1.2) (see Chapter 1) is replaced by

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i) \tag{2.4}$$

with subscript i the time index. Each i corresponds to a half-hour time slot. $\delta\mathbf{x}$ is as before the increment at low resolution at initial time, and $\delta\mathbf{x}(t_i)$ the increment evolved according to the tangent linear model from the initial time to time index i. $\mathbf{R}_i$ and $\mathbf{B}$ are the covariance matrices of observation errors at time index i and of background errors respectively. $\mathbf{H}_i$ is a suitable linear approximation at time index i of the observation operator H_i. The innovation vector is given at each time step by $\mathbf{d}_i = \mathbf{y}_i^{\mathrm{o}} - H_i\mathbf{x}^{\mathrm{b}}(t_i)$, where $\mathbf{x}^{\mathrm{b}}(t_i)$ is the background propagated in time using the full non-linear model and $\mathbf{y}_i^{\mathrm{o}}$ is the observation vector at time index i. If the switch LTC=.true. (default is LTC=.false.), then SYNOP and DRIBU time sequences of surface pressure and height data are used with serial correlation of observation error, and the observation cost-function computation for those data then spans all time slots. Equation (2.4) therefore needs generalising, as has been done in the paper by Järvinen *et al.* (1999).

The minimization is performed in the same way as in 3D-Var. However, it works fully in terms of increments, a configuration which is activated by the switches L131Tl and LOBSTL, and involves running the tangent-linear and adjoint models iteratively as explained in Section 2.4 of Chapter 2, and using the tangent-linear observation operators.

A way to account in the final 4D-Var analysis for some non-linearities is to define a series of minimization problems such that

$$\begin{aligned} J(\delta\mathbf{x}^n) = {} & \frac{1}{2}(\delta\mathbf{x}^n + \mathbf{x}^{n-1} - \mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x}^n + \mathbf{x}^{n-1} - \mathbf{x}^{\mathrm{b}}) \\ & + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\delta\mathbf{x}^n(t_i) - \mathbf{d}_i^{n-1})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}^n(t_i) - \mathbf{d}_i^{n-1}) \end{aligned} \tag{2.5}$$

with superscript n the minimization index.

$\mathbf{x}^{n-1}$ is the current estimate of the atmospheric flow. It is equal to the background for the first minimization. $\mathbf{d}_i^{n-1} = \mathbf{y}_i^{o} - H_i\mathbf{x}^{n-1}(t_i)$ is the innovation vector, computed by integrating the model at high resolution from the current estimate.

2.6 TANGENT LINEAR PHYSICS

The linearized physics describes five processes: vertical diffusion, subgrid-scale orographic effects, radiation, large-scale condensation and deep convection. In order to prevent spurious unstable perturbations from growing, a number of simplifications have been defined for these schemes with respect to the nonlinear physical parametrization schemes (described in Part IV: Physical processes) used in the forecast model.

The first minimization has been using just a simple physics, which is represented by very simple vertical diffusion scheme acting mainly close to the surface and surface drag scheme. A scientific description of this simple physics is given in Buizza (1994). From Cy28r3 onwards the more detailed simplified vertical diffusion scheme described in subsection Subsection 2.6.3 is being used instead of Buizza's in the first minimization.

The following minimizations use a more complete linearized physics, which is described in this section. The text is focused on brief scientific description of these linearized physical parametrization schemes and some technical aspects. More detailed scientific issues can be found elsewhere (Mahfouf *et al.*, 1997; Rabier *et al.*, 1997; Mahfouf, 1999; Janisková *et al.*, 2002).

2.6.1 Set-up

The simple physics used in the first minimization is activated by the switches LSPHLC, LVDFDS, LSDRDS, LVDFLC, LSDRLC, LKEXP in namelist NAPHLC.

In order to activate the improved linear physics, the switch LSPHLC of the simplified linear physics in NAPHLC should be set to .FALSE.. In CVA1 when both logicals LSPHLC and LETRAJP are equal to .TRUE., LSPHLC is reset to .FALSE. and a warning is written in the standard output (logical unit NULOUT).

The improved linearized physics is activated by the switches LETRAJP, LEVDIF2, LEGWDG2, LECOND2, LERADI2, LERADS2, LERADN2, LERADSW2, LEDCLD2, LECUMF2 in namelist NAMTRAJP. The following switches must be set to .TRUE. – LEPHYS, LAGPHY (also necessary to activate the ECMWF nonlinear physics) in namelist NAEPHY and LETRAJP (to activate storage of the trajectory at $t - \Delta t$). The linear physics contains a set of five physical processes: vertical diffusion (LEVDIF2), subgrid-scale orographic effects (LEGWD2), large scale condensation (LECOND2), radiation (LERADI2, LERADS2, LERADN2, LERADSW2) taking into account clouds (LEDCLD2) and deep moist convection (LECUMF2).

Tunable parameters of the improved physics (which should not in principle be modified) are defined in SUPHLI. The logical LPHYLIN is used to activate the simplifications and/or modifications associated with the linear package in the nonlinear physics. This variable is set to .FALSE. by default, but is forced to .TRUE. before starting the tangent-linear and adjoint computations in CNT3TL and CNT3AD. Thus this switch remains .TRUE. for the linear physics called in EC_PHYS_TL and EC_PHYS_AD.

2.6.2 Mixed-phase thermodynamics

The thermodynamical properties of the water mixed phase are represented by a differentiable weighting function between $T_0 = 0°\mathrm{C}$ and $T_{\mathrm{ice}} = -23°\mathrm{C}$ given by

$$\alpha(T) = \frac{1}{2}[1 + \tanh\{\mu(T - T_{\mathrm{crit}})\}] \tag{2.6}$$

with $\mu = 0.15$ (RLPALP1) and $T_{\mathrm{crit}} = T_{\mathrm{ice}} + \frac{T_0 - T_{\mathrm{ice}}}{\sqrt{2}}$ (RLPTRC).

The tuning parameter μ controls the intensity of the smoothing, and the temperature T_{crit} has been chosen to give $\alpha = 0.5$ for the same temperature as in the operational quadratic formulation (see function FCTTRE).

This weighting function is used by the large-scale condensation and moist-convection routines.

2.6.3 Vertical diffusion

Vertical diffusion applies on wind components, dry static energy and specific humidity. The exchange coefficients in the planetary boundary layer and the drag coefficients in the surface layer are expressed as functions of the local Richardson number (Louis *et al.*, 1982). They differ from the operational formulation which uses the Monin–Obukhov length as a stability parameter in stable conditions and a K-profile approach for convective boundary layers (see the documentation of the ECMWF physics). Analytical expressions are generalized to the situation of different roughness lengths for heat and momentum transfer. The mixing length profile $l(z)$ uses the formulation of Blackadar (1962) with a reduction in the free atmosphere.

For any conservative variable ψ (wind components u, v; dry static energy s; specific humidity q), the tendency of its perturbation ψ' produced by vertical diffusion is

$$\frac{\partial \psi'}{\partial t} = \frac{1}{\rho}\frac{\partial}{\partial z}\left(K(Ri)\frac{\partial \psi'}{\partial z}\right) \tag{2.7}$$

where ρ is the air density.

In the planetary boundary layer, the exchange coefficient K is given by

$$K = l^2 \left\| \frac{\partial \mathbf{V}}{\partial z} \right\| f(Ri) \tag{2.8}$$

where $f(Ri)$ represents the coefficient of the vertical turbulent diffusion which is a function of the Richardson number. The mixing length vertical profile is defined as

$$l_{\mathrm{MH}} = \frac{\kappa(z+z_0)}{1+k\frac{(z+z_0)}{\lambda}}\left[\gamma + \frac{1-\gamma}{1+\frac{(z+z_0)^2}{L^2}}\right] \tag{2.9}$$

The asymptotic mixing length λ_{M} for momentum is set to 150 m, whereas $\lambda_{\mathrm{H}} = \lambda_{\mathrm{M}}\sqrt{1.5d}$. The pseudo-depth of the boundary layer is defined by $L = 4$ km (RLPMIXL), and the reduction factor applied to the mixing length in the free atmosphere is $\gamma = 0.2$ (RLPBETA) [$1 \rightarrow \gamma\lambda$ when $z \gg L$]. κ is von Kármán's constant and z_0 is the roughness length.

The coefficients of the vertical turbulent diffusion for momentum are given by

$$f(Ri) = \begin{cases} \dfrac{1}{1+\frac{2bRi}{\sqrt{1+dRi}}} & \text{if } Ri > 0, \text{ the stable case} \\ 1 - \dfrac{2bRi}{1+3bc\left(\frac{1}{\sqrt{27}}\right)\left(\frac{l_{\mathrm{M}}}{z+z_{0\mathrm{M}}}\right)^2\sqrt{|Ri|}} & \text{if } Ri < 0, \text{ the unstable case} \end{cases} \tag{2.10}$$

The vertical diffusion coefficients for dry static energy and specific humidity are defined as

$$f(Ri) = \begin{cases} \dfrac{1}{1+3bRi\sqrt{1+dRi}} & \text{if } Ri > 0 \\ 1 - \dfrac{3bRi}{1+3bc\left(\frac{1}{\sqrt{27}}\right)\left(\frac{l_{\mathrm{H}}}{z+z_{0\mathrm{H}}}\right)^2\sqrt{|Ri|}} & \text{if } Ri < 0 \end{cases} \tag{2.11}$$

where b, c, d are constants controlling the dependency of vertical exchange on the stability.

To parametrize the turbulent fluxes at the surface, the drag coefficients (i.e. the exchange coefficients between the surface and the lowest model level) are computed. The neutral coefficients C_{MN} and C_{HN} are written as

$$C_{\mathrm{MN}} = \frac{\kappa^2}{\left[\ln\left(\frac{z+z_{0\mathrm{M}}}{z_{0\mathrm{M}}}\right)\right]^2} \quad \text{and} \quad C_{\mathrm{HN}} = \frac{\kappa^2}{\left[\ln\left(\frac{z+z_{0\mathrm{M}}}{z_{0\mathrm{M}}}\right)\ln\left(\frac{z+z_{0\mathrm{M}}}{z_{0\mathrm{H}}}\right)\right]} \tag{2.12}$$

The drag coefficients for momentum are computed as

$$C_{\mathrm{M}} = \begin{cases} \dfrac{1}{1+\frac{2bRi}{\sqrt{1+dRi}}}C_{\mathrm{MN}} & \text{if } Ri > 0 \\ \left(1 - \dfrac{2bRi}{1+3bcC_{\mathrm{MN}}\sqrt{\frac{z+z_{0\mathrm{M}}}{z_{0\mathrm{M}}}|Ri|}}\right)C_{\mathrm{MN}} & \text{if } Ri < 0 \end{cases} \tag{2.13}$$

The surface exchange coefficients for dry static energy and specific humidity are

$$C_{\mathrm{H}} = \begin{cases} \dfrac{1}{1+3bRi\sqrt{1+dRi}}C_{\mathrm{HN}} & \text{if } Ri > 0 \\ \left(1 - \dfrac{3bRi}{1+3bcC_{\mathrm{HN}}\sqrt{\frac{z+z_{0\mathrm{M}}}{z_{0\mathrm{M}}}|Ri|}}\right)C_{\mathrm{HN}} & \text{if } Ri < 0 \end{cases} \tag{2.14}$$

The empirical coefficients b (RLPBB), c (RLPCC) and d (RLPDD) are set to 5 in SUPHLI.

In (2.7) perturbations of the exchange coefficients are neglected, $K' = 0$, in order to prevent spurious unstable perturbations from growing in the linearized version of the scheme (Mahfouf, 1999). This is also the case for the exchange coefficients between the surface and the lowest model level.

This modified diffusion scheme is called directly from VDFMAIN and makes use of all the routines from the operational vertical diffusion, except VDFSFLX (computation of the surface fluxes for tiles) and some diagnostic routines. However, the exchange coefficients are computed in a different way in VDFEXCS and VDFEXCU.

The linear versions of the vertical diffusion scheme are called from the drivers VDFMAINTL and VDFMAINAD. The linearization of the surface energy balance is also performed (VDFTSKTL, VDFTSKAD), but perturbations of the skin temperature are not evolved in time (in EC_PHYS_TL and EC_PHYS_AD). This simplification should be relaxed when the skin temperature becomes part of the control variable. There are tangent-linear and adjoint versions of some routines, which were not updated when introducing the surface tile scheme and are not recently used. This is secured by the logical switch LENOPERT set to .TRUE. in SUPHLI (i.e. no perturbation is required for the surface arrays). This concerns the routines for computation of: the roughness length (VDFUPDZ0TL, VDFUPDZ0AD), the surface boundary conditions for T and q (VDFSURFTL, VDFSURFAD), the surface exchange coefficients (VDFEXCSTL, VDFEXCSAD) and the equivalent evapotranspiration efficiency (VDFEVAPTL, VDFEVAPAD).

The logical LEKPERT in NAMTRAJP controls the perturbations of the exchange and drag coefficients. When set to .FALSE. (default), the perturbations are set to 0. to prevent the growth of spurious instabilities in the tangent-linear model.

2.6.4 Subgrid scale orographic effects

Only the low-level blocking part of the operational nonlinear scheme developed by Lott and Miller (1997) (see documentation on ECMWF nonlinear physics) is taken into account in the tangent-linear and adjoint calculations. The deflection of the low-level flow around orographic obstacles is supposed to occur below an altitude $Z_{\rm blk}$ such that

$$\int_{Z_{\rm blk}}^{2\mu} \frac{N}{|\mathbf{U}|}\,{\rm d}z \geq H_{n_{\rm crit}} \tag{2.15}$$

where $H_{n_{\rm crit}}$ is a critical non-dimensional mountain height (GFRCRIT = 0.5), μ is the standard deviation of subgrid-scale orography, N is the vertical stability and $\mathbf{U}$ is the wind vector.

The deceleration of the wind due to the low-level blocking is given by

$$\begin{aligned}\left(\frac{\partial \mathbf{U}}{\partial t}\right)_{\rm blk} &= -C_{\rm d}\max\left(2-\frac{1}{r},0\right)\frac{\sigma}{2\mu}\left(\frac{Z_{\rm blk}-z}{z+\mu}\right)^{0.5}(B\cos^2\psi + C\sin^2\psi)\frac{\mathbf{U}|\mathbf{U}|}{2}\\ &= A(\mathbf{U}|\mathbf{U}|)\end{aligned} \tag{2.16}$$

where

$C_{\rm d}$ is the low-level drag coefficient (GWAKE = 1)
σ is the mean slope of the subgrid-scale orography
γ is the anisotropy of the subgrid-scale orograph
ψ is the angle between the low-level wind and the principal axis of topography
$B = 1 - 0.18\gamma - 0.04\gamma^2$
$C = 0.48\gamma + 0.3\gamma^2$
$r = (\cos^2\psi + \gamma\sin^2\psi)/(\gamma\cos^2\psi + \sin^2\psi)$

Since the final wind tendency produced by the low-level drag parametrization is obtained from the following partially implicit discretization of (2.16)

$$\left(\frac{\partial \mathbf{U}}{\partial t}\right)_{\rm blk} = -\frac{\beta}{\beta+1}\frac{\mathbf{U}^{n-1}}{2\Delta t} \tag{2.17}$$

where $\beta = A|\mathbf{U}^{n-1}|2\Delta t$, the corresponding perturbation is

$$\left(\frac{\partial \mathbf{U}}{\partial t}\right)_{\rm blk} = -\frac{\beta'}{(\beta+1)^2}\frac{\mathbf{U}^{n-1}}{2\Delta t} - \frac{\beta}{\beta+1}\frac{\mathbf{U}'^{n-1}}{2\Delta t} \tag{2.18}$$

Finally, the local dissipation heating is computed in the same way as in the nonlinear scheme, together with the associated perturbation.

The main tangent-linear and adjoint routines, GWDRAGTL and GWDRAGAD, are called from CALLPARTL and CALLPARAD respectively. The depth of the low-level blocking layer is determined in GWSETUPTL and GWSETUPAD, while the low-level blocking effect described by (2.17) is computed at the end of GWDRAGTL and GWDRAGAD. As mentioned above, the representation of wave breaking is currently not activated in the linearized code by setting the constant RLPDRAG to zero in SUPHLI. Note that RLPDRAG is only used when logical LPHYLIN is .TRUE., otherwise GKDRAG is used (set to 0.3 in SUGWD).

2.6.5 Radiation

The radiation scheme solves the radiative transfer equation in two distinct spectral regions. The computations for the longwave (LW) radiation are performed over the spectrum from 0 to 2820 cm^{-1}. The shortwave (SW) part of the scheme integrates the fluxes over the whole shortwave spectrum between 0.2 and 4.0 μm. The scheme used for data assimilation purposes must be computationally efficient to be called at each time step and at the full spatial resolution for an improved description of the cloud-radiation interactions during the assimilation period (Janisková *et al.*, 2002).

(*a*) *The short-wave radiation scheme*

The linearized code for the shortwave radiation scheme has been derived from the ECMWF original nonlinear scheme developed by Fouquart and Bonnel (1980) and revised by Morcrette (1991). In this scheme, the photon-path-distribution method is used to separate the parametrization of the scattering processes from that of molecular absorption. Upward $F_{\mathrm{sw}}^{\uparrow}$ and downward $F_{\mathrm{sw}}^{\downarrow}$ fluxes at a given level j are obtained from the reflectance and transmittance of the atmospheric layers as

$$F_{\mathrm{sw}}^{\downarrow}(j) = F_0 \prod_{k=j}^{N} T_{\mathrm{bot}}(k) \tag{2.19}$$

$$F_{\mathrm{sw}}^{\uparrow}(j) = F_{\mathrm{sw}}^{\downarrow}(j) R_{\mathrm{top}}(j-1) \tag{2.20}$$

Computations of the transmittance at the bottom of a layer T_{bot} start at the top of atmosphere and work downward. Those of the reflectance at the top of the same layer R_{top} start at the surface and work upward. In the presence of cloud in the layer, the final fluxes are computed as a weighted average of the fluxes in the clear sky and in the cloudy fractions of the column as

$$R_{\mathrm{top}} = C_{\mathrm{cloud}} R_{\mathrm{cloud}} + (1 - C_{\mathrm{cloud}}) R_{\mathrm{clear}} \tag{2.21}$$

$$T_{\mathrm{top}} = C_{\mathrm{cloud}} T_{\mathrm{cloud}} + (1 - C_{\mathrm{cloud}}) T_{\mathrm{clear}} \tag{2.22}$$

In the previous equations, C_{cloud} is the cloud fractional coverage of the layer within the cloudy fraction of the column (depending on cloud-overlap assumption).

The cloud optical properties depend on three different parameters: the optical thickness, the asymmetry factor and the single scattering albedo. These parameters are derived from Fouquart (1987) for water clouds, and Ebert and Curry (1992) for ice clouds. They are functions of cloud condensate and a specified effective radius. All cloudy fluxes are computed from cloud optical thicknesses derived from the liquid and ice cloud water contents weighted by a 0.7 inhomogeneity factor following Tiedtke (1996).

The nonlinear version of the shortwave radiation scheme (described in details in Part IV: Physical processes) uses six spectral intervals with transmission functions derived from a line-by-line code (Dubuisson *et al.*, 1996).

The nonlinear scheme is reasonably fast for application in 4D-Var and has, therefore, been linearized without *a priori* modifications. The only modification with respect to the nonlinear model is using two spectral intervals with transmission functions (instead of six intervals used in the nonlinear model) in order to reduce the computational cost.

(*b*) *The longwave radiation scheme*

The longwave radiation scheme, operational in the ECMWF forecast model until June 2000, was a band emissivity type scheme (Morcrette, 1989). This scheme has been replaced by the Rapid Radiation Transfer Model (RRTM: Mlawer *et al.*, 1997) in June 2000. The scheme is based on a two-stream solution of the longwave transfer. The ECMWF version of the RRTM longwave radiation scheme (Morcrette, 1998) also includes a maximum-random overlap assumption but keeps the cloud fraction and cloud optical thickness as two separate quantities. In RRTM, the cloud optical thickness is defined as a function of spectrally varying mass absorption coefficients and relevant cloud water and ice paths. This quantity is then used within the actual cloudy fraction of the layer.

Cloud longwave optical properties are represented by the emissivity $\varepsilon_{\rm cld}$ related to the condensed water amount, and by the condensed water mass absorption coefficient $k_{\rm abs}$ obtained following Smith and Shi (1992) for water clouds and Ebert and Curry (1992) for ice clouds.

The complexity of the RRTM radiation scheme for the longwave part of the spectrum makes accurate computations expensive. In the variational assimilation framework, simplifications were made to reduce its computational cost. A combination of artificial neural networks (Chevallier *et al.*, 2000) and pre-computed Jacobians of the longwave radiation flux with respect to global mean temperature and water vapour profiles (Chevallier and Mahfouf, 2001) has been defined for the linearized longwave radiation scheme.

The longwave radiative fluxes depend upon temperature, water vapour, cloud cover and liquid and ice water contents. The design of the scheme allows the separation of the contribution of temperature and water vapour from that of cloud parameters. More precisely, the upward and downward longwave fluxes at a certain height z_i are expressed as

$$F(z_i) = \sum_k a_k(z_i) F_k(z_i) \tag{2.23}$$

where the coefficients a_k are a function of the cloudiness matrices $CC_{i,k}$ computed using some overlap assumption. Differentiating the above equation, a flux perturbation is computed as

$$\mathrm{d}F(z_i) = \sum [\underbrace{a_k(z_i)}_{\text{(a)}} \underbrace{\mathrm{d}F_K(z_i)}_{\text{(b)}} + \underbrace{f_K(z_i)}_{\text{(c)}} \underbrace{\mathrm{d}a_k(z_i)}_{\text{(d)}}] \tag{2.24}$$

(a) NL model, (b) Jacobian matrices, (c) NeuroFlux, (d) TL model

In the proposed approach, the coefficients ak are computed using the nonlinear (NL) model (code computing cloud optical properties). Due to the weak nonlinearities in the variations of the radiative fluxes F_k with respect to temperature and water vapour, the tangent-linear approximation can be used to compute perturbations of radiation fluxes $\mathrm{d}F(z_k)$ from pre-computed mean Jacobian matrices. Perturbations of radiative fluxes with respect to cloud parameters $\mathrm{d}a_k(z_i)$ are computed using a tangent-linear (TL) scheme. The trajectory of radiative fluxes required in the tangent-linear and adjoint computations are efficiently estimated from a neutral network version of the ECMWF longwave radiative transfer model (called NeuroFlux) in order to avoid significant extra memory storage or costly re-computations. The Neutroflux is significantly faster (seven times) than the operational longwave radiation code with a comparable accuracy.

(*c*) *Cloud overlap assumptions*

Cloud overlap assumptions must be made in atmospheric models in order to organize the cloud distribution used for radiation and precipitation/evaporation computations. A cloud overlap assumption of some sort is necessary to account for the fact that clouds often do not fill the whole grid box. The maximum-random overlap assumption (originally introduced in Geleyn and Hollingsworth, 1997) is used operationally in the ECMWF model (Morcrette, 2000). Adjacent layers containing cloud are combined by using maximum overlap to form a contiguous cloud and discrete layers separated by clear-sky are

combined randomly as

$$CC_{i,j} = 1 - (1 - C_i) \prod_{k=i+1}^{j-1} \left[\frac{1 - \max(C_k, C_{k-1})}{1 - C_{k-1}} \right] \tag{2.25}$$

where $CC_{i,j}$ is cloudiness encountered between any levels i and j in the atmosphere and C_k is the cloud fraction of the layer k located between levels k and $k+1$.

The tangent-linear (TL) and adjoint (AD) versions of the shortwave radiation scheme can be activated using the logical switch LERADSW2 = .TRUE. in the namelist NAMTRAJP. The default value is set to .FALSE.. To decrease a computational cost of the shortwave radiation in data assimilation, the number of spectral intervals is reduced to two (six intervals are used in the forecast model) during minimization. The number of spectral intervals NSW is changed in CVA1. This change also requires to read several parameters for two spectral intervals. This is done in set-up routines SUSWN, SUCLOPN and SUAERSN.

Due to the low impact of cloud perturbations produced by the current linearized cloud scheme and in order to decrease the computational cost, a reduced version of the linearized longwave radiation (i.e. $\mathrm{d}F(z_i) \approx \sum_k a_k(z_i)\, \mathrm{d}F_k(z_i)$) is currently used in data assimilation. This reduced version can be activated by LERADN2 = .TRUE. in the namelist NAMTRAJP. If the full linearized longwave radiation should be used, the logical switch LERADFL2 must also be set to .TRUE. in the same namelist. The default values are set to .FALSE..

The linearized radiation schemes are called from RADINATL and RADINAAD, where the computation of radiation fluxes is performed. Tendencies produced by the linearized longwave and shortwave radiation are computed in RADHEATTL and RADHEATAD. All those routines are called from CALLPARTL and CALLPARAD, respectively.

RADLSWTL and RADLSWAD are the drivers for the computation of the solar and thermal fluxes by calling specialized routines for shortwave radiation (SWTL and SWAD) and for longwave radiation (LWPTL and LWPAD).

2.6.6 Cloud scheme

The ECMWF diagnostic cloud scheme (Slingo, 1997), used before the implementation of the current operational prognostic scheme (Tiedtke, 1993), has been linearized to be used with the linearized radiation scheme.

The diagnostic cloud scheme allows for four cloud types: convective cloud and three types of layer clouds (high, middle and low level). The convective clouds (C_{conv}) are parametrized using the scale-averaged precipitation rate ($\bar{P}$) from the model convection scheme as

$$C_{\mathrm{conv}} = 0.4 \min(0.8, 0.125 \ln \bar{P} - 1.5) \tag{2.26}$$

They can fill any number of model layers and their depth is determined by the convection scheme. The stratiform clouds (C_{strat}) are determined from a function of the layer relative humidity (RH_e) after adjustment for the presence of convective clouds, $RH_e = RH - C_{\mathrm{conv}}$, as

$$C_{\mathrm{strat}} = f(RH_e) = \left\{ \max\left(\left[\frac{RH_{\mathrm{e}} - RH_{\mathrm{crit}}}{1 - RH_{\mathrm{crit}}}, 0 \right] \right) \right\}^2 \tag{2.27}$$

RH_{crit} is a critical relative humidity for cloud formation defined as

$$RH_{\mathrm{crit}} = 1 - \max[0, 1.8\sigma(1-\sigma)(1 + 0.95(\sigma - 0.5))] \tag{2.28}$$

where σ is the vertical coordinate of the pressure divided by the surface pressure.

Liquid water l_{lwc} and ice water l_{iwc} are proportional to the specific humidity at saturation q_{sat}. They are defined as

$$l_{\mathrm{lwc}} = \alpha l_{\mathrm{lc}} \tag{2.29}$$

$$l_{\mathrm{iwc}} = (1 - \alpha) l_{\mathrm{lc}} \tag{2.30}$$

where α is the fraction of condensate held as liquid water and $l_{\rm lc} = 0.5 C_{\rm cloud} q_{\rm sat}$ with $C_{\rm cloud}$ being cloud cover.

The linearized diagnostic cloud scheme has been further simplified by accounting for part of the dependencies only. The current linearized convection scheme does not provide perturbations of precipitation. This leads to a zero perturbation of convective cloudiness since it is determined from the convective precipitation rate. A modification of the linearized cloud scheme was also necessary in order to avoid the threshold problem in the parametrization of the layer clouds computed as a function of relative humidity $RH_{\rm e}$ (see (2.27)). For this reason, the derivative of the function of relative humidity has been set to zero (Janisková *et al.*, 2000).

The linearized versions of the diagnostic cloud scheme are CLOUDTL and CLOUDAD and they are called from CALLPARTL and CALLPARAD, respectively. These linearized schemes can be activated using the logical switch LEDCLD2 = .TRUE. in the namelist NAMTRAJP. The default value is set to .FALSE..

2.6.7 Large-scale condensation

The linearized version of large-scale condensation scheme is based on a local moist adjustment scheme designed to produce stratiform precipitation as described in Mahfouf (1999).

If (q, T) is an initial supersaturated state $(q > q_{\rm s}(T))$, a final adjustment state (q_*, T_*) such that $q_* = q_{\rm s}(T_*)$ is reached by conserving moist static energy in a given atmospheric layer. The adjusted values are obtained through an iterative method with two updates.

The equations for the nonlinear scheme are

$$T_* - T = -\lambda(q_{\rm s}(T_*) - q) \tag{2.31}$$

$$q_* - q = q_{\rm s}(T_*) - q \tag{2.32}$$

with $\lambda = L/c_p$, where L is the latent heat of liquid water and c_p is the specific heat at constant pressure. They are solved using a linearization of the specific humidity at saturation: $q_{\rm s}(T_*) = q_{\rm s}(T) + \gamma(T_* - T)$ with $\gamma = \partial q_{\rm s}/\partial T$. Then, the linearization of $q_{\rm s}(T')$ as $\gamma T'$, provides the equations for the tangent-linear scheme so that

$$T'_* - T' = -\frac{\lambda}{1+\lambda\gamma}(\gamma T' - q') \tag{2.33}$$

$$q'_* - q' = -\frac{\lambda}{1+\lambda\gamma}(\gamma T' - q') \tag{2.34}$$

Tendencies produced by large-scale condensation are thus

$$\frac{\partial q'}{\partial t} = -\frac{1}{\lambda}\frac{\partial T'}{\partial t} = \frac{q'_* - q'}{2\Delta t} \tag{2.35}$$

where Δt is the model time step.

This scheme also accounts for melting of snow whenever the temperature of a given layer exceeds 2\dg C. Evaporation of precipitation in subsaturated layer is neglected to avoid the growth of spurious unstable modes Mahfouf (1999).

Linearized versions of the large-scale condensation scheme are CONDTL and CONDAD. Local supersaturation is removed through a local moist-adjustment scheme (CUADJTQTL, CUADJTQAD). Supersaturation produces precipitation instantaneously (without a cloud stage). The effect of rainfall evaporation in sub-saturated layers is disregarded in the linearized versions of the scheme by setting the constant RLPEVAP to 0., instead of 0.95 in the nonlinear parametrization.

2.6.8 Deep moist convection

According to Betts (1997) the physical tendencies produced by convection on any conservative variable ψ (wind components, specific humidity, dry static energy) can be written in a mass-flux formulation as

$$\frac{\partial \psi}{\partial t} = \frac{1}{\rho}\left[(M_{\rm up} + M_{\rm down})\frac{\partial \psi}{\partial z} + D_{\rm up}(\psi_{\rm up} - \psi) + D_{\rm down}(\psi_{\rm down} - \psi)\right] \tag{2.36}$$

The first term on the right-hand side represents the compensating subsidence induced by cumulus convection on the environment through the mass flux M. The other terms accounts for the detrainment of cloud properties in the environment with a detrainment rate D. The subscripts '*up*' and '*down*' refer to the updrafts and downdrafts properties respectively. Evaporation of cloud water and precipitation should also be added in (2.36) for the dry static energy s and specific humidity q. Mass-flux theory predicts that the effect of convection on large-scale temperature and moisture structures is dominated by compensating subsidence, as shown in the study of Gregory and Miller (1989). Therefore, nonlinear tendencies produced by convection can be approximated by

$$\frac{\partial \psi}{\partial t} \approx \frac{1}{\rho}\left[(M_{\text{up}} + M_{\text{down}})\frac{\partial \psi}{\partial z}\right] \tag{2.37}$$

This approach is better for the heat budget than for the moisture budget (Betts, 1997). As a consequence, a partial linearization of (2.37) is performed, where the vertical transport by the mean mass flux is applied to the perturbations. ¿From a nonlinear integration of the convection scheme, mass-flux profiles $M(z)$ are used to estimate the vertical transport of perturbed fields ψ' according to

$$\frac{\partial \psi'}{\partial t} = \frac{1}{\rho}\left[(M_{\text{up}} + M_{\text{down}})\frac{\partial \psi'}{\partial z}\right] \tag{2.38}$$

This linearized equation only applies to deep convection. The underlying assumption of (2.38) is that perturbations of cloud properties can be neglected.

The mass-flux profiles associated with the updrafts M_{up} and the downdrafts M_{down} are recomputed in the tangent-linear and adjoint integrations from the stored basic state. This partial linearization implies that all the nonlinear routines for the convection scheme have their tangent-linear and adjoint counterparts (starting form the driving routines CUCALLNTL and CUCALLNAD). However, most of them are only used to recompute the trajectory. The only routines which contain linear statements are CUININTL (mean thermodynamical properties at half model levels), CUDUDVTL (tendencies for the wind components) and CUDTDQNTL (tendencies for dry static energy and specific humidity). Equation (2.38) is solved numerically in the following form (see Tiedtke, 1989) as

$$\frac{\partial \psi'}{\partial t} = \frac{1}{\rho}\left[\frac{\partial (M_{\text{up}}\psi')}{\partial z} + \frac{\partial (M_{\text{down}}\psi')}{\partial z} + (D_{\text{up}} - E_{\text{up}} + D_{\text{down}} - E_{\text{down}})\psi'\right] \tag{2.39}$$

which requires extra local storage of the profiles of entrainment E and detrainment D rates computed in CUASCN and in CUDDRAFN (variables PDMFEN and PDMFDE). Equation (2.39) is only applied when deep convection is diagnosed from the basic state.

2.6.9 Trajectory management

The ECMWF physics uses the tendencies from the dynamics and variables at $t - \Delta t$ as input to compute the tendencies of a given process (represented by the operator P) for a prognostic variable ψ. Therefore

$$\frac{\psi^{n+1} - \psi_{\text{u}}^{n-1}}{2\Delta t} = \text{P}(\psi_{\text{u}}^{n-1}) \tag{2.40}$$

where the variable ψ_{u} has already been updated by the dynamics and by the previous physical processes (which are called in the following order: radiation; vertical diffusion; subgrid-scale orographic effects; moist convection; large-scale condensation).

Thus

$$\psi_{\text{u}}^{n-1} = \psi^{n-1} + \left(\frac{\partial \psi}{\partial t}\right)_{\text{dyn}} + \left(\frac{\partial \psi}{\partial t}\right)_{\text{phys}} \tag{2.41}$$

In (2.40), if the operator P is nonlinear, its linearization around the basic state ψ_{u}^{n-1}, will require to store the model state at time step $n-1$ (trajectory at $t - \Delta t$) as well as the tendencies produced by the dynamics $(\partial\psi/\partial t)_{\text{dyn}}$. The physical tendencies from the previous processes $(\partial\psi/\partial t)_{\text{phys}}$, require an additional call to the nonlinear routines in the adjoint computations (CALLPARAD) and a local storage of the partial tendencies.

The storage of the trajectory at $t - \Delta t$ is performed in EC_PHYS by the routine STORE_TRAJ_PHYS called before the driver of the ECMWF physics CALLPAR. Fields are stored in grid-point space in an array TRAJ_PHYS. This array is allocated in the module TRAJ_PHYSICS, where also the number of the fields to be stored is defined.

The following three-dimensional fields are stored.

(i) **For the atmosphere:** the prognostic variables (wind components, temperature, specific humidity) and their tendencies produced by adiabatic processes, the vertical velocity, the long-wave fluxes and the solar transmissivity.
(ii) **For the soil:** the prognostic variables for temperature and moisture content (used to compute the surface fluxes from the trajectory in the linear vertical-diffusion scheme).
(iii) **For the tiles (i.e. vegetation (surface cover) types):** u- and v-stress, surface sensible heat flux, surface evaporation and skin temperature.

A number of two-dimensional fields used at time step $t - \Delta t$ need to be stored: surface pressure, surface fluxes, skin temperature, skin reservoir, snow reservoir, roughness lengths (mainly for the vertical diffusion).

The preliminary computations (pressure and geopotential at full and half model levels, astronomy parameters) are performed in EC_PHYS_TL and EC_PHYS_AD before calling the driver of the tangent-linear physics CALLPARTL or the driver of the adjoint physics CALLPARAD, and after reading the trajectory fields from GET_TRAJ_PHYS.

Chapter 3

3D variational assimilation (with the first-guess at appropriate time)

Table of contents

3.1 INTRODUCTION

3D-Var is a temporal simplification of 4D-Var. The simplest way to implement a 3D-Var within the context of an existing 4D-Var scheme is to replace the tangent-linear (and its adjoint) model integration within the inner loops (as defined in the previous section) by the identity operator (LIDMODEL = .TRUE. in yomtnewt.F90). This approach has indeed been adopted, as it saves on maintenance of scripts and codes. It is this 3D-Var version that was used for the production of the ERA-40 re-analysis (Uppala *et al.*, 2005) and it was used until 14 March 2006 in ECMWF operations within the so called BC-suite to generate timely boundary conditions for the participating member states.

In this version of 3D-Var as much as possible of the 4D-Var structure is maintained. The available observations, typically over the period of a 6-hour assimilation window, are compared with a model integration at high resolution. The comparison between observation and model is thus performed at the appropriate time: therefore the abbreviation 3D-FGAT (first-guess at appropriate time). This configuration is activated via the switch LFGAT = .TRUE. in namvar.h.

3.2 3D-FGAT

3.2.1 Solution algorithm

The tangent-linear model in 4D-Var evolves the analysis increment over time, within the assimilation window. In 3D-FGAT no such evolution takes place: one static increment in produced. The valid time of this increment is not at the initial time of the window as in 4D-Var, but at the central time. The increment is added to the background field, which forms the analysis at the central time. This has the consequence that a trajectory update from the start of the assimilation window cannot be performed. The 3D-FGAT configuration is therefore based on one single outer-loop without iteration. There are nevertheless two minimisation steps. The first provides analysis error and background error estimates required for cycling, and also pre-conditioning vectors (Hessian eigenvectors) that are used in the second minimisation job-step. The resolution of the first minimisation is typically lower (T42) than for the main one (e.g. T159).

Scatterometer de-aliasing is active and variational quality control is activated approximately halfway into the minimisation, rendering the problem non-linear (and the cost function non-quadratic). The use of conjugate-gradient minimisation is thus prevented. The M1QN3 algorithm is used instead.

3.2.2 Main differences with respect to 4D-Var

The main differences with respect to the standard operational 4D-Var configuration are as follows.

(i) Variational quality control is carried out within the inner-loop iterations (LQVARQC = .FALSE.).
(ii) The scatterometer de-aliasing is performed within the inner-loop iterations (LQSCAT = .FALSE.).
(iii) Use of M1QN3 instead of conjugate gradient minimisation, due to the cost-function being quadratic.
(iv) The J_c term relies on NMI (normal mode initialisation) rather than DFI (Digital filter initialisation) (LJCDFI = .FALSE., LJCNMI = .TRUE.).
(v) No iteration at outer-loop level.
(vi) LTRAJHR = .FALSE., which means that the observation operators are linearized around the low-resolution trajectory, which is generated through an integration by the full non-linear model (at the resolution of the minimisation) at the beginning of the minimisation job-step. CVA1 calls CNT2 to do this.

3.2.3 Data selection

The observational data are stored in time-slots within the ODB (just as in 4D-Var). The data selection criteria are applied in identically the same way as in 4D-Var, allowing time sequences of data to be used also in 3D-FGAT. The 3D scheme is however unable to extract temporal information from the data, and produces an analysis of the temporal average of the FGAT departures. Optionally the screening data selection rules can be applied once for the entire (6-hour) assimilation window, which would pick only the data closes to the analysis centre time. This is called '3D-screening' and was the practice in the original operational 3D-Var configuration.

Chapter 4
Background term

Table of contents

4.1 INTRODUCTION

The background term described in Courtier *et al.* (1998) was replaced in May 1997 by a new formulation by Bouttier *et al.* (1997), and replaced again in April 2005 by a wavelet-based covariance model (Fisher (2004), Fisher (2003)). The two older formulations are still part of the IFS but will not be described in this documentation.

4.2 DESCRIPTION OF THE ALGORITHM

We use the following notation.

(i) $\delta\mathbf{x}$ is the low-resolution analysis increment (i.e. model field departures from the background).
(ii) $\mathbf{B}$ is the assumed background error covariance matrix.
(iii) ζ, η, (T, p_{surf}) are increments of vorticity, divergence, temperature and surface pressure on model levels.
(iv) η_{b} and $(T, p_{\mathrm{surf}})_{\mathrm{bal}}$ are the *balanced* parts of the η and (T, p_{surf}) increments. The concept of balance will be defined below.
(v) η_u and $(T, p_{\mathrm{surf}})_{\mathrm{unbal}}$ are the *unbalanced* parts of η and (T, p_{surf}), i.e. η–η_{bal} and $[(T, p_{\mathrm{surf}})$–$(T, p_{\mathrm{surf}})_{\mathrm{bal}}]$, respectively.

The incremental variational analysis problem, (1.2) of Chapter 1, is rewritten in the space defined by the change of variable $\delta\mathbf{x} = \mathbf{L}\chi$ (Section 1.4) where $\mathbf{L}$ satisfies $\mathbf{L}\mathbf{L}^{\mathrm{T}} = \mathbf{B}$ so that J_{b} takes a simple form. In operational practice, the initial point of the minimization is either the background (in which case $\delta\mathbf{x} = \chi = 0$) or the control vector saved from an earlier minimization. The minimization is carried out in the space of χ, where J_{b} is the euclidean inner product. At the end of the minimization, the analysis increments are reconstructed in model space by $\delta\mathbf{x} = \mathbf{L}\chi$. In order to compare with observations $\mathbf{x}$ is reconstructed using (2.4), in each simulation. Thus the variational analysis can be performed using only transformations from minimization space to model space (CHAVARIN). The transformation from model space to minimization space is never required. In particular, $\mathbf{L}$ is not required to be invertible, and may even be rectangular. This is the case in the "wavelet" J_{b} formulation described here, where the dimension of the control vector is considerably larger than the dimension of the model state vector.

The background-error covariance matrix $\mathbf{B}$ is implied by the design of $\mathbf{L}$, which currently has the form

$$\mathbf{L} = \mathbf{K}\mathbf{L}_{\mathrm{u}} \tag{4.1}$$

where $\mathbf{K}$ is a balance operator going from the set of variables ζ, η_{u}, $(T, p_{\mathrm{surf}})_{\mathrm{u}}$, (etc.) to the model variables ζ, η, (T, p_{surf}), (etc.). The $\mathbf{L}_{\mathrm{u}}$ operator defines the covariance matrix for ζ, η_{u}, $(T, p_{\mathrm{surf}})_{\mathrm{u}}$ (etc.), as

$$\mathbf{B}_{\mathrm{u}} = \mathbf{L}_{\mathrm{u}}\mathbf{L}_{\mathrm{u}}^{\mathrm{T}} \tag{4.2}$$

So far, the formulation is perfectly general. Now, we restrict $\mathbf{L}_{\mathrm{u}}$ to a simple form and choose a particular balance operator $\mathbf{K}$.

The covariance matrix $\mathbf{B}_{\mathrm{u}}$ is assumed to be block-diagonal, with no correlation between the parameters, so that

$$\mathbf{B}_{\mathrm{u}} = \begin{bmatrix} \mathbf{C}_{\zeta} & 0 & 0 & 0 \\ 0 & \mathbf{C}_{\eta_{\mathrm{u}}} & 0 & 0 \\ 0 & 0 & \mathbf{C}_{(T,p_{\mathrm{surf}})_{\mathrm{u}}} & 0 \\ 0 & 0 & 0 & \ddots \end{bmatrix} \tag{4.3}$$

The matrix $\mathbf{L}_{\mathrm{u}}$ is similarly block-diagonal, with diagonal blocks $\mathbf{L}_{\zeta}$, $\mathbf{L}_{\eta_{\mathrm{u}}}$, $\mathbf{L}_{(T,p_{\mathrm{surf}})_{\mathrm{u}}}$, (etc.). However, these sub-matrices are rectangular. Each sub-matrix is treated identically in the code, except for differences in the coefficents used to describe the covariances, so we will consider just one sub-matrix, $\mathbf{L}_{\zeta}$.

4.2.1 Wavelet J_{b}

The "wavelet" J_{b} formulation was devised to allow both spatial and spectral variation of the horizontal and vertical covariances of background error. Only a brief description is given here. The reader is referred to Fisher (2004) for a mathematical justification of the method, and also to Fisher (2003).

Simultaneous spatial and spectral variation of horizontal and vertical covariances is achieved by dividing the control vector up into several parts, each of which corresponds to a band of total spherical wavenumbers, n. For each band, the elements of the control vector are arranged on a linear reduced Gaussian grid that is appropriate to the spectral truncation corresponding to the highest wavenumber in the band. The wavenumber bands overlap, with the result that exactly two bands use linear grids corresponding to the full model resolution. The cutoff wavenumber decreases by approximately a factor of $\sqrt{2}$ for each subsequent band and the number of gridpoints reduces by a factor of two, so that the total number of elements of the control vector is approximately three times larger than a gridpoint representation of the model fields.

Information about the vertical and horizontal correlations is stored in the form of sets of vertical covariance matrices. There is one set of matrices for each wavenumber band, and for each band the symmetric square-roots of these matrices are stored on a horizontal grid that is somewhat coarser than the grid used for the band's control vector elements.

The transformation of the control vector to model space, represented by $\mathbf{L}_{\zeta}$, consists of the following steps. First, for each wavenumber band, each vertical column of the grid is multiplied by the square-root of the covariance matrix (from the set corresponding to the wavenumber band) that is nearest to the grid column. Next, the control vector elements are transformed to spectral space. Since the wavenumber bands overlap, there are now more than one (in fact, exactly two) sets of spectral coefficients for each spherical wavenumber, n. These sets of coefficients are combined in a weighted sum to give a conventional spectral description of the model fields. For each wavenumber, the sums of squares of the weights is equal to one. In fact, the weights are the square-roots of triangle functions that decay to zero towards the boundaries of each wavenumber band, and take the value one at the centre of the band. As a result, the vertical covariance associated with each wavenumber is effectively linearly interpolated between the covariance matrices defined for each band, and the horizontal covariance is defined by a piecewise-linear function of n.

The final step of the transformation $\mathbf{L}_{\zeta}$ is to transform the spectral fields to the model grid, and to multiply by the assumed standard deviations of background error.

4.2.2 The balance operator

The balance relationship takes the following form

$$\begin{aligned}(T, p_{\text{surf}})_{\text{b}} &= \mathbf{N}\zeta + \mathbf{P}\eta_{\text{u}} \\ \eta_{\text{b}} &= \mathbf{M}\zeta + \mathbf{Q}_1(T, p_{\text{surf}}) + \mathbf{Q}_2\zeta\end{aligned} \tag{4.4}$$

The matrix blocks $\mathbf{M}$, $\mathbf{N}$, $\mathbf{P}$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are in general not invertible, but the balance operator$\mathbf{K}$ is. The $\mathbf{M}$, $\mathbf{N}$ and $\mathbf{P}$ operators used to define balance have a restricted algebraic structure. $\mathbf{M}$ and $\mathbf{N}$ are both the product of a so-called horizontal balance operator H by vertical balance operators M, N such that

$$\begin{aligned}\mathbf{M} &= MH \\ \mathbf{N} &= NH\end{aligned} \tag{4.5}$$

The H operator is a block–diagonal matrix of identical horizontal operators transforming the spectral coefficients of vorticity, independently at each level, into an intermediate variable P_{b} which is a kind of linearized mass variable. The horizontal operators in H are defined analytically as linearized versions of the non-linear balance equation:

$$\nabla^2 P_{\text{b}} = (f + \zeta) \times v_\psi + \frac{1}{2}\nabla(v_\psi \cdot v_\psi) \tag{4.6}$$

where $v_\psi = \mathbf{k} \times \nabla\psi$ is the rotational wind.

This equation is simplified, by treating model levels as if they were pressure levels, and is linearised about the background state to provide a linear equation for increments in P_{b} as a function of vorticity increments.

The matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are also defined analytically, as simplified and linearised versions of the quasi-geostrophic omega equation:

$$\left(\sigma\nabla^2 + f_0^2\frac{\partial^2}{\partial p^2}\right)\omega' = -2\nabla \cdot \mathbf{Q} \tag{4.7}$$

(Here, $\mathbf{Q}$ is the Hoskins' Q-vector: a function of temperature and rotational wind.)

Once again, model levels are treated as if they were pressure levels, and the stability parameter, σ is assumed to be a function of pressure. This allows the equation to be separated into a set of small tri-diagonal systems that can be solved non-iteratively to give divergence increments as a linear function of vorticity and temperature increments.

The M, N and $\mathbf{P}$ operators all have the same structure: block-diagonal, with one full vertical matrix per spectral component. The vertical matrices depend only on the total wavenumber n, and are calculated by linear regression. (For example, in the case of the operator N, the regression is between $\mathbf{P}$ and (T, p_{surf}).)

Calibration of the statistically-derived operators of the balance operator is performed outside the IFS using programs developped by Francois Bouttier.

4.3 IMPLEMENTATION

4.3.1 The J_{b} information structure: SPJB_VARS_INFO

The J_{b} code is required to be sufficiently flexible to allow the incorporation of new variables, and to allow the choice of analysed variables to differ from the choice of model variables. This is achieved through a separation of the arrays used by J_{b} from the model arrays, and is controlled by a "structure" (Fortran derived type) containing information about the J_{b} variables: SPJB_VARS_INFO. This structure is ubiquitous throughout the J_{b} code.

SPJB_VARS_INFO is defined in YOMJG, and is a one-dimensional array with one element for each analysed variable. Analysed variables may be either three-dimensional (for example, vorticity) or two-dimensional (e.g. surface pressure).

Each element of SPJB_VARS_INFO is a structure with the following elements:

- IGRIBCODE
- IGRIBCODE_FCE
- IPT
- IPTJB
- IPTFCE
- IPTBG
- L_IN_SPA3
- L_IN_SPA2
- L_IN_SPGFL
- L_IN_GPGFL
- COR_STRING

The first two elements give the GRIB parameter codes used to identify the variable and the corresponding background error fields. These may be different since, for fields involved in the balance operator, IGRIBCODE_FCE refers to the unbalanced component of the field, whereas IGRIBCODE refers to the full (balanced plus unbalanced) field.

The elements IPT, IPTJB, IPTFCE and IPTBG are indexes into various arrays. IPT indicates which field in the model's SPA3, SPA2, SPGFL or GFL arrays corresponds to the analysed variable. IPTJB provides an index into the (spectral) J_b arrays, SPA3JB and SPA2JB. IPTFCE locates the variable in the forecast-error array FCEBUF, and IPTBG locates the corresponding background fields in SP7A3, GP7A3 or SP7A2.

Four logical flags, L_IN_SPA3, L_IN_SPA2, L_IN_SPGFL and L_IN_GPGFL, show in which of the model's various arrays the variable is stored. These flags are used in conjuntion with the element IPT to locate the model's equivalent of the analysed variable.

The final element of SPJB_VARS_INFO is a character string. This is used to match the variable with the covariance matrices in the input file "wavelet.cv".

4.3.2 User input

The initial setup of J_b is performed by SUJB. The user can input, via NAMJG, the number of analysed variables (N_SPJB_VARS). GRIB parameter codes for these variables (M_GRIBCODES) and for the corresponding background error fields (M_GRIBCODES_FCE), and also the character strings that identify the corresponding covariance matrices (C_COR_STRINGS), are input via NAMJBCODES.

These inputs are sufficient for SUJB to construct the information structure, SPJB_VARS_INFO. The correspondence between analysis fields and model fields is derived by looking up the grib code of the analysed variable in the model's GMV and GFL structures. If none of the inputs is specified, the analysis defaults to 6 variables: vorticity, divergence, temperature, surface log-pressure, humidity and ozone.

Background error statistics for wavelet J_b may be calculated by setting LJBWSTATS to TRUE. In this case, the code expects to read a large set of background states as spectral GRIB fields. The states are expected to be produced by an analysis ensemble with N_BGMEMBERS members and covering N_BGDATES dates. Inter-member differences are calculated for each date, and the statistics of background error are calculated from this set of differences. (Statistics may be calculated using the "NMC method" by regarding, for example, a set of 48-hour forecasts as member 1, and a set of 24-hour forecasts as member 2.) The filename for each state is constructed as a six-character prefix (CINBGSTATES) followed by a string of the form "mXXX_dYYY", where XXX is in the range 1...N_BGMEMBERS, and YYY is in the range 1...N_BGDATES.

Calculation of the statistics is performed in SUJBWAVGEN and associated subroutines. At the end of the calculation, a file called "wavelet.cv" is written, and the code stops a little inelegantly with an abort message. Note that calculation of the J_b statistics is largely single-threaded.

4.3.3 Input files

The IFS requires three input files to define J_b. The covariance matrices that define the vertical and horizontal correlations are read from "wavelet.cv". This file also contains power spectrum information which is used to determine global mean vertical profiles of standard deviations of background error. Coefficients for the balance operator are read from "stabal96.bal", and standard deviations of background error are read from "errgrib".

4.3.4 Initial setup

SUJBWAVELET0 determines the number of wavenumber bands and their wavenumber boundaries by reading header information from the statistics file "wavelet.cv". Next, the spherical transforms and grid layouts corresponding to the cutoff wavenumbers for each band are defined by a call to SUJBWAVTRANS. This information is stored in the structure GRID_DEFINITION (YOMWAVELET).

Some allocation of arrays is carried out in SUALGES, and the control vector is allocated in SUALCTV. Coefficients for the humidity variable and for the balance operators are initialised by SUJBCHVAR and SUJBBAL, respectively.

Most of the setup for "wavelet J_b" is performed by SUJBWAVELET and SUJBWAVALLO. The main tasks are to read the input file, which contains eigenvalues and eigenvectors of vertical covariance matrices; to calculate the symmetric square-roots of these matrices; and to distribute the matrices to those processors that will need them. In addition, the input file contains power-spectrum informtion about the analysed variables, which is used to construct global-mean profiles of background error standard deviation.

The covariance matrices are stored in the structure WAVELET_VCORS, together with information about their grid layout.

The final part of the J_b setup is to read the background fields (SUECGES), and to initialise the standard deviations of background error (SUINFCE).

4.3.5 Change of variable

The change of variable, CHAVARIN, implicitly defines the background error covariance model. It consists of three steps: CVAR2IN implements multiplication of control vector by the matrix denoted $\mathbf{L}$ in Section 4.2. The transformed variables are inserted into the model's arrays by JBTOMODEL. This may involve spectral transforms, since the J_b variables are all spectral, whereas model variables may be held as gridpoint fields. Finally, the background state is added to the model fields by ADDBGS.

The control vector for the minimization contains several components that are unrelated to J_b. These are dealt with in CVAR2IN. The J_b components are dealt with by a call to SQRTBIN. This, in turn, calls CVARU3I to apply the covariance model, and then applies the balance operator via calls to BALNONLINTL, BALVERT and BALOMEGATL.

CVARU3I calls JGCORI to apply the vertical and horizontal correlations, and CVARGPTL to multiply by standard deviations of background error. The latter routine calls TRANSINV_WAVELET and TRANSDIR_WAVELET to perform the spectral transforms required to convert variables between spectral and gridpoint representations. Note that these routines implement *spectral* transforms using the information stored in GRID_DEFINITION. They do not implement wavelet transforms.

JGCORI calls JBVCOR_WAVELETIN once for each band of wavenumbers to perform the main part of the wavelet J_b change of variable. JBVCOR_WAVELETIN has three steps. First, each column of the control vector is multiplied by the square-root of a vertical covariance matrix. This is performed by JBVCORG. Next, the gridpoint fields are transformed to spectral space by a call to TRANSDIR_WAVELET. The coefficients are multiplied by the appropriate weights by WAVXFORM, and the weighted coefficients are added into the arrays PSPA3JB and PSPA2JB.

The calls to JBVCORG are the most expensive part of the wavelet J_b code. The calls are performed within an Open-MP loop over NPROMA blocks. For each grid column and each analysis variable, the nearest covariance matrix is identified by a call to JBMATINTERP, and the column is multiplied by

the square-root of the covariance matrix. To optimize this calculation, each matrix has an associated "skyline", which identifies the first and last element in each row (or column, since the matrices are symmetric) that is significantly different from zero. Only the matrix elements between these limits are included in the multiplication, resulting in a considerable increase in computational efficiency.

Chapter 5

Observation operators and the observation cost function (J_o)

Table of contents

5.1 INTRODUCTION

The observation operators provide the link between the analysis variables and the observations (Lorenc, 1986; Pailleux, 1990). The observation operator is applied to components of the model state to obtain the model equivalent of the observation, so that the model and observation can be compared. The operator H in (1.4) signifies the ensemble of operators transforming the control variable $\mathbf{x}$ into the equivalent of each observed quantity, $\mathbf{y}^o$, at observation locations. The 3D/4D-Var implementation allows H to be (weakly) non-linear, which is an advantage for the use of satellite radiance data, scatterometer data with aliased wind direction, cloud and precipitation data, for example. In this chapter we define the content of each of the IFS observation operators and the associated controls for all data used in 3D/4D-Var. The calculation of departures and the J_o costfunction are also described.

The IFS observation operators are generic in the sense that the same routine is often used for several different data types. For example, the radiance operator (RTTOV) simulates measurements from a large number of satellite radiometers (microwave and infrared), and the temperature operator (PPT) is used for TEMP, AIREP, and other data types and it is also used to provide input to RTTOV. Similarly the routine PPQ is used for interpolation of specific humidity to given pressures, but it can also be used for any other atmospheric mixing ratio constituents, such as ozone and carbon dioxide. Note that many of the PP-routines were developed for the model's pressure-level post-processing package and are used also in that context.

In Table 10.3 there is a list of the observing systems currently used by 4D-Var in ECMWF's operational data assimilation suite. The table also indicates important restrictions on data usage and thinning of data. The observation errors are also given in Chapter 10. In 4D-Var, the evolving model state is compared to the available observations at the correct time, currently at half-hourly intervals. These intervals are called *time-slots*. The Observation Data Base (see ODB documentation) holds the observations organized by time slots. If there are multiple reports from the same fixed observing station within a time slot then the data nearest the analysis time are selected for use in the analysis. Some thinning is applied for satellite data and other moving platforms reporting frequently. The thinning rules are applied to each time slot, separately. The thinning, quality control and data selection tasks are performed in the *screening* job step of 4D-Var – it is described in Chapter 10.

5.1.1 Data selection controls, and the interface to the blacklist

Most data selection criteria are coded in so called *blacklist files*, written in a convenient, readable blacklist language (see the Blacklist Documentation, Järvinen *et al.*, 1996). The blacklist mechanism is very flexible and allows nearly complete control of which data to use/not use in the assimilation. The 'monthly blacklist' is the part of the blacklist that is based on operational data monitoring results, and it is maintained by the Meteorological Operations Section. The blacklist is consulted in the screening job. The interface is set up in BLINIT, in such a way that a number of named items from the header (Table 5.1) and body (Table 5.2) parts of the observation report can be passed to the blacklist software. Depending on the blacklisting criteria flags are communicated to the routine BLACK, and those are written to the ECMA ODB data base. Blacklist-rejected data are subsequently excluded from the CCMA ODB and will not be present in the minimisation job steps. Data selection rules should be coded in the blacklist files whenever possible rather than in the IFS code itself. The operational blacklist history is kept in an archive.

Classes of data can also be switched on and off using the NOTVAR array in NAMJO, however it is preferable to use the blacklist mechanism for this purpose. The second dimension in this array is the observation type. The first dimension is the variable number (NVAR, see below). The elements of the NOTVAR array can take either of two values: 0, means that the data will be used; -1, means that the data will not be used.

5.2 HORIZONTAL INTERPOLATION TO OBSERVATION POINTS

5.2.1 Method

Currently it is assumed that each observation equivalent can be computed from a single vertical profile of model data. That is, it is assumed that each observation operator can be written as $H = H_v H_h(\mathbf{x})$ where

Table 5.1 *Header variables for the ifs/blacklist interface.*

Index	Name	Description
1	OBSTYP	observation type
2	STATID	station identifier
3	CODTYP	code type
4	INSTRM	instrument type
5	DATE	date
6	TIME	time
7	LAT	latitude
8	LON	longitude
9	STALT	station altitude
10	LINE_SAT	line number (atovs)
11	RETR_TYP	retrieval type
12	QI_1	quality indicator 1
13	QI_2	quality indicator 2
14	QI_3	quality indicator 3
15	MODORO	model orography
16	LSMASK	land-sea mask (integer)
17	RLSMASK	land-sea mask (real)
18	MODPS	model surface pressure
19	MODTS	model surface temperature
20	MODT2M	model 2m temperature
21	MODTOP	model top level pressure
22	SENSOR	satellite sensor indicator
23	FOV	field of view index
24	SATZA	satellite zenith angle
25	NANDAT	analysis date
26	NANTIM	analysis time
27	SOE	solar elevation
28	QR	quality of retrieval
29	CLC	cloud cover
30	CP	cloud top pressure
31	PT	product type
32	SONDE_TYPE	sonde type

Table 5.2 *Body variables for the ifs/blacklist interface.*

Index	Name	Description
1	VARIAB	variable name
2	VERT_CO	type of vertical coordinate
3	PRESS	pressure, height or channel number
4	PRESS_RL	reference level pressure
5	PPCODE	synop pressure code
6	OBS_VALUE	observed value
7	FG_DEPARTURE	first guess departure
8	OBS-ERROR	observation error
9	FG_ERROR	first-guess error
10	WINCHAN_DEP	window channel departure
11	OBS_T	observed temperature

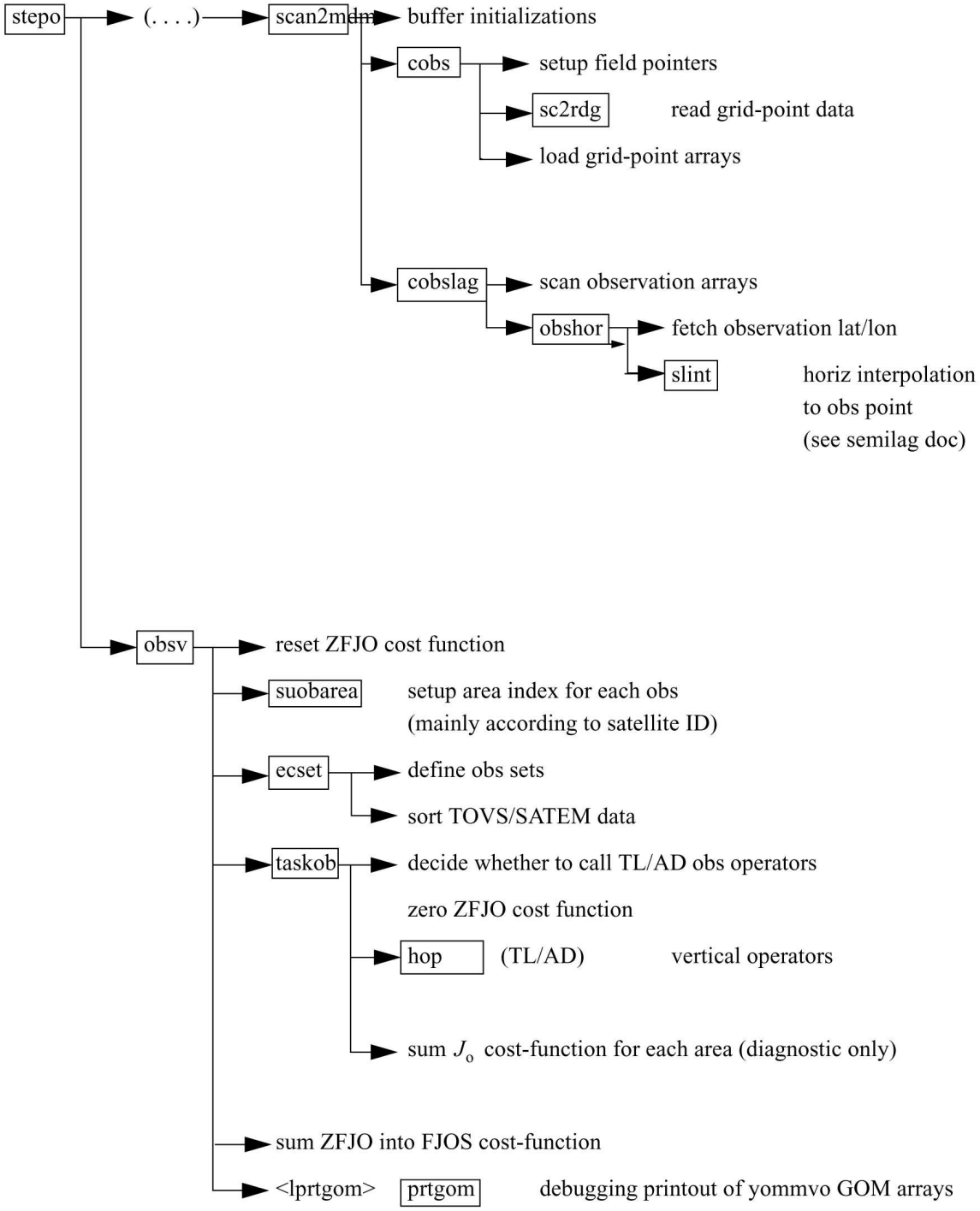

Figure 5.1 *Flow diagram for subroutines scan2mdm and obsv.*

H_h is horizontal interpolation of model data to the location of the observation. H_h is performed for all observations within a time-slot directly after the corresponding model time step has been performed. The output of H_h is a vertical profile of model data for each observation location. These profiles are stored in a data structure call GOM-arrays, and used later in the vertical part of the observation operator H_v performed in HOP, called from OBSV. Recently, 2D GOM-arrays have been developed, to permit horizontal integration as required for limb-sounding and radio-occultation data. The orientation of the 2D GOM-arrays (e.g. one vertical plane per observation) is then determined by the satellite viewing geometry. The H_v operator for those data receive the vertical plane of model data as input.

COBSLAG is the master routine for the horizontal interpolation of model data to observation points. It is called after the inverse spectral transform in SCAN2MDM, and after the so-called *semi-Lagrangian buffers* have been prepared by COBS and SLCOMM2, see the flow diagram in Fig. 5.1. The interpolation code is shared with the semi-Lagrangian advection scheme of the dynamics. The buffers contain a *halo* of gridpoints big enough to enable interpolation to all observations within the grid-point domain belonging to the processor. COBSLAG calls OBSHOR which does the following.

(i) Performs the interpolation, using SLINT.
(ii) Message-passes the result to the processors where the corresponding observations belong, and copies the model data at observation points to the so-called GOM-arrays (GOMS, described below) using the routine MPOBSEQ.

There are three methods of horizontal interpolation.

(i) LAIDDIOBS: 12-point bi-cubic interpolation, used for all upper-air fields (if NOBSHOR = 203) except clouds.
(ii) LAIDDIOBS: Bi-linear interpolation, used for surface fields.
(iii) LAIDLIC: Nearest gridpoint, used for cloud parameters.

The interpolation method for the upper-air fields can be switched to bi-linear by specifying NOBSHOR = 201 in namobs. The default is NOBSHOR = 203 (bi-cubic) for the trajectory jobs, but NOBSHOR = 201 (bi-linear) in the minimisation jobs. Lists of interpolation points and weights are prepared by the routine LASCAW. In 4D-Var bi-cubic interpolation is used at high resolution (i.e. in the trajectory runs), and bi-linear is used at low resolution (i.e. in the main minimization). The interpolation is invoked once per 4D-Var time slot.

The adjoint (OBSHORAD) follows the same general pattern but gets further complicated by the fact that the gradient from several observations may contribute to the gradient at a given gridpoint. The summation of gradients is to done in the same order, irrespective of the number of processors, as reproducibility is desired (strictly speaking, overall 4D-Var reproducibility is only ensured if the so called reproducibility switch is ON. This adds to computational cost, and the default is OFF). Tables for managing this are prepared in MKGLOBSTAB.

5.2.2 Storage in GOM-arrays

There are two types of GOM arrays, one-dimensional GOM arrays which contain vertical profiles of model values at observation points and two-dimensional GOM arrays which contain a set of vertical profiles of model values. There are namelist-switches (in NAMGOM) to control which upper-air model variables that will appear in the GOM-arrays and whether to use one or two dimenional GOM arrays. The control is at at the observation type level.

The setup of the GOM structures are done in routine SUGOMS and the data structures can all be found in module GOMS

The trajectory GOM5 arrays (identical to GOM) are allocated in the case that tangent linear observation operators are used. They are to hold the trajectory interpolated to the observation locations, and in that case the GOM-arrays hold the perturbations.

At the end of the adjoint observation operators the GOM-arrays are zeroed and overwritten by the gradient (in PREINTAD).

The r.m.s. of the GOM arrays is printed (by PRTGOM) if the switch LPRTGOM = .TRUE., (in YOMOBS). The default is that the print is switched on. It can be located in the log file by searching for 'RMS OF GOM'. The printing is done from OBSV, (i) when the GOM arrays contain the background interpolated to the observation points, (ii) when it contains ∇J_o of the first simulation, (iii) when it contains first TL perturbations after the initial call to the minimizer and (iv) when it contains ∇J_o at the final simulation.

5.2.3 Variable numbers and association with observation operators

In the ODB each observed value is associated with a vertical position ('press' with pointer MDBPPP, given in terms of pressure, height or channel number) and a variable number ('varno' with pointer MDBVNM). The defined variable numbers are listed in the array NVNUMB as described in the ODB documentation. The variable number indicates which physical quantity has been observed. Each observed quantity is linked to an IFS observation operator (in HVNMTLT). In the case there is no corresponding observation operator in IFS, no observation equivalent will be computed, and the observation will be rejected from further processing. Each available observation operator has been given a number (NVAR) and a short name (CVAR_NAME, three characters), set in YOMCOSJO.

There are currently JPXVAR = 25 operator names defined: U, U10, DD, FF, H, H2, T, Z, DZ, LH, T2, TS, RAD, SN, RR, PS, CC, CLW, Q, FFF, S0, X, PWC, TO3 and TCW, numbered sequentially in NVAR. Each number can be referenced by variables such as NVAR_U(= 1), NVAR_U10(= 2) and so on. Based on the ODB variable number for each observation, which has been associated with an observation-operator number and name, HOP calls the appropriate observation operator routine. For example, observations with ODB variablenumber = NVNUMB(8), will in HVNMTLT be associated with NVAR_T = 7 and CVAR_NAME(7) = 'T', and HOP will thus call the subroutine PPT.

In the log file CVAR_NAME is used to label entries in the J_o-table. The J_o-table shows the J_o-values for each observation sub-type, for each observation operator (NVAR). RMS of observation error and background error are also shown. See Subsection 5.2.6.

5.2.4 Organization in observation sets

The vertical observation operators are vectorized over NMXLEN (yomdimo) data. To achieve this the data first have to be sorted by type and subdivided into sets of lengths not exceeding that number. NMXLEN is currently set to 511, in SUDIMO. The observation sets may span several 4D-Var time slots, as the input to the observation operators is the GOM-arrays which have been pre-prepared for all time slots during the tangent linear model integration. However, in the current operational context this is not the case, as OBSV.F90 is called once per time slot. Furthermore, on the current IBM machines long vector loops are not very important for performance, our IBMs being scalar machines.

The organization of the sets is done in ECSET and SMTOV and the information about the sets is kept in yomobset. Radiance observation sets (e.g. ATOVS, AIRS, SSMI) must not contain data from more than one instrument. That is to say that inside HOP, HOPTL and HOPAD (and routines called form there) data from several instruments are not mixed – different instrument are handled in separate calls. This is controlled by sorting according to the area-parameter, which for radiance data is an indicator of satellite ID and instrument, prior to forming the sets. The area-parameter is determined in SUOBAREA, and is irrelevant for the observation processing for all data other than the radiance data.

5.2.5 Cost function

The master routine controlling the calls to the individual observation operators is called HOP. This routine deals with all different types of observations.

The HOP/HOPTL/HOPAD routines are called from TASKOB/TASKOBTL/TASKOBAD (called from OBSV/OBSVTL/OBSVAD) in a loop over observation sets. The data type of each set is know from the information in tables such as MTYPOB(KSET) stored in yomobset.

The following describes HOP//HOPTL/. The adjoint HOPAD follows the reverse order.

(i) First prepare for vertical interpolation using the routine PREINT. Data on model levels are extracted from the GOM-arrays (GOMS). Pressures of model levels are computed using GPPRE. Help arrays for the vertical interpolation are obtained (PPINIT) and T^* and T_0 are computed (CTSTAR). T^* and T_0 are later used for extrapolation of temperature below the model's orography, Subsection 5.5.1. The routine PREINTS deals with model surface fields needed for the near-surface observation operators and PREINTR deals with those fields that are specific to the radiance observation operators.

Table 5.3 *Association between variable numbering and observation operator routines. The CVAR-NAMEs also appear in the printed J_o-table in the log-file.*

NVAR	CVAR_NAME	Observation operator routine	Description
1	U	PPUV	Upper air wind components
2	U10	PPUV10M	10-metre wind components
3	DD		Wind direction
4	FF	PPUV	Wind speed
5	H	PPRH	Relative humidity
6	H2	PPRH2M	2-metre relative humidity
7	T	PPT	Temperature
8	Z	PPGEOP	Geopotential
9	DZ	PPGEOP	Thickness
10	LH	PPRH	Layer mean RH (Meteo-France)
11	T2	PPT2M	2-metre temperature
12	TS		Surface temperature (M-France)
13	RAD	RADTR	Radiance data
14	SN		Snow (M-France)
15	RR		Rain rate (Meteo-France)
16	PS	PPPS	Surface pressure
17	CC	PPTCC	Cloud cover
18	CLW	PPCLW	Cloud liquid water
19	Q	PPQ	Specific humidity
20	FFF	PPUV10M	10-metre wind speed
21	S0		Sigma_0
22	X		Reserved
23	PWC	PPPWC	Layer water content or TCWV
24	TO3	PPPWC	Layer ozone content
25	TCW		Layer cloud water content
26	RFL	REFLSIM	Radar reflectivity
27	APD	GPSZEN_DELAY	GPS total zenith delay
28	RO	GPSRO_OP	GPS radio occultation
29	HLS	RTL_HOP_1D	Limb sounding radiance
30	AOD	AOD_OP	Aerosol optical depth
31	LRA		unknown

(ii) The observation array is then searched to see what data is there. The 'body' of each observation report is scanned for data, and the vertical coordinate and the variable-number for each datum is retained in tables (ZVERTP and IVNMRQ). These tables will later constitute the 'request' for model equivalents to be computed by the various observation operators. Tables of pointers to data ('body' start addresses) and counters are stored (arrays IPOS and ICMBDY).

(iii) Then the forward calculations are performed. There is an outer loop over all known 'variable numbers'. If there are any matching occurrences of the loop-variable number with the content of IVNMRQ, then the relevant observation operator will be called. A variable-number and an observation operator are linked by a table set up in the routine HVNMTLT. The interface routines PPOBSA (upperair) and PPOBSAS (surface) are used, which in turn call PPFLEV and the individual operator routines. For radiance data the interface is RADTR which calls the radiative transfer code.

(iv) In HDEPART, calculate the departure $\mathbf{z}$ as

$$\mathbf{z} = \mathbf{y}^o - H\mathbf{x} + (\mathbf{y}^o - H\mathbf{x}^{\mathrm{b}}_{\mathrm{HR}}) - (\mathbf{y}^o - H\mathbf{x}^{b}_{\mathrm{LR}}) \tag{5.1}$$

where the two terms in brackets have been computed previously: the first one in the high resolution trajectory run and the second one in the LOBSREF call, described in Section 2.2.

If LOBSTL then $\mathbf{z}$ is

$$\mathbf{z} = \mathbf{y}^o - \mathbf{H}\delta\mathbf{x} + (\mathbf{y}^o - H\mathbf{x}^b_{\rm HR}) - \mathbf{y}^o \tag{5.2}$$

which simplifies to what has been presented in Section 1.4.
The TOVS radiance bias correction is also carried out at this point by subtracting the bias estimate (kept in the NCMTORB-word of ODB) from the calculated departure.
Finally the departure is divided by the observation error σ_o (NCMFOE in ODB) to form the *normalized departure*.

(v) Departures of correlated data are multiplied by $\mathbf{R}^{-1}$, see Subsection 5.2.7. The division by σ_o has already taken place in HDEPART, so $\mathbf{R}$ at this point is in fact a correlation (not a covariance) matrix.

(vi) The cost function is computed in HJO, as

$$J_o = \mathbf{z}^{\rm T}\mathbf{z} \tag{5.3}$$

for all data, even for SCAT data when LQSCAT = .TRUE., as in current 4D-Var with quadratic cost-function. When LQSCAT = .FALSE. (as in current 3D-Var configuration) the SCAT cost function combines the two ambiguous winds (subscripts 1 and 2) in the following way (also in HJO),

$$J_{\rm SCAT} = \left[\frac{J_1^4 J_2^4}{J_1^4 + J_2^4}\right]^{1/4} \tag{5.4}$$

These expressions for the cost function are modified by variational quality control, see Section 5.3.
The cost-function values are store in two tables, as detailed in Subsection 5.2.6.

(vii) HJO, also stores the resulting *effective departure* in the NCMIOM0-word of ODB, for reuse as the input to the adjoint. The effective departure is the normalized departure after the effects of (vertical) observation error correlation and quality control have been taken into account, $\mathbf{z}_{\rm eff} = \mathbf{z}^{\rm T}\mathrm{R}^{-1}[\mathbf{QC}_{\rm weight}]$, where the QC-weight will be defined below, Section 5.3 .

We have now reached the end of the forward operators. In the adjoint routine HOPAD some of the tasks listed above have to be repeated before the actual adjoint calculations can begin. The input to the adjoint (the effective departure) is read from the ODB. The expression for the gradient (with respect to the observed quantity) is then simply

$$\nabla_{\rm obs} J_o = -2\mathbf{z}_{\rm eff}/\sigma_o \tag{5.5}$$

which is calculated in HOPAD for all data. The gradient of $J_{\rm SCAT}$ is much more complicated and is calculated in a separate section of HOPAD. The adjoint code closely follows the structure of the direct code, with the adjoint operators applied in the reverse order.

5.2.6 J_o tables

There are two different tables for storing the J_o values. One is purely diagnostic (JOT, yomcosjo.h), and is used for producing the printed J_o tables in the log-file (PRTJO called rom EVCOST). JOT is a FORTRAN90 derived type with items containing cost function value, data count, observation error, background error and codetype, for each observation type and for each observed variable (as defined by NVAR above).

The actual J_o-table is called FJOS (yomcosjo). FJOS is indexed by the absolute observation number, iabnob = MABNOB(jobs, kset), so that the contributions from each individual observation can be summed up in a predetermined order (in EVCOST), to ensure reproducibility, irrespective of number of processors and tasks.

5.2.7 Correlation of observation error

The observation error is assumed uncorrelated (i.e. the matrix $\mathbf{R}$ is diagonal) for all data. Optionally, by setting LTC=.true., time-sequences of SYNOP/DRIBU surface pressure and height data (Järvinen *et al.*, 1999) will be time correlated. There used to be code for vertical correlation of observation error for radiosonde geopotential data and SATEM thicknesses, but these were removed in Cy21r2. Similar can

easily be reintroduced again in a later cycle, provided the correlation acts on data within a single report (vertical or inter-channel correlation).

The serial correlation for SYNOP and DRIBU data is modelled by a continuous correlation function $ae^{-b(t_1-t_2)^2}$ where $a = \text{RTCPART} = 0.3$ and $b = \text{RTCEFT} = 6.0$ hours, under the switch LTC (namjo). The remaining fraction $1-a$ of the error variance is assumed uncorrelated (see COMTC).

When $\mathbf{R}$ is non-diagonal, the 'effective departure' $\mathbf{z}_{\text{eff}}$ is calculated by solving the linear system of equations $\mathbf{z}_{\text{eff}}\mathbf{R}$ for $\mathbf{z}_{\text{eff}}$, using LAPACK routines SPOTRF (Choleski decomposition) and SPOTRS (backwards substitution), as is done in COMTC.

5.3 VARIATIONAL QUALITY CONTROL

The variational quality control, VarQC, has been described by Andersson and Järvinen (1999). It is a quality control mechanism which is incorporated within the variational analysis itself. A modification of the observation cost function to take into account the non-Gaussian nature of gross errors, has the effect of reducing the analysis weight given to data with large departures from the current iterand (or preliminary analysis). Data are not irrevocably rejected, but can regain influence on the analysis during later iterations if supported by surrounding data. VarQC is a type of buddy check, in that it rejects those data that have not been fitted by the preliminary analysis (i.e. the current state vector), often because it conflicts with surrounding data.

5.3.1 Description of the method

The method is based on Bayesian formalism. First, an *a priori* estimate of the probability of gross error $P(G)_{\text{i}}$ is assigned to each datum, based on study of historical data. Then, at each iteration of the variational scheme, an *a posteriori* estimate of the probability of gross error $P(G)_{\text{f}}$ is calculated (Ingleby and Lorenc, 1993), given the current value of the iterand (the preliminary analysis). VarQC modifies the gradient (of the observation cost function with respect to the observed quantity) by the factor $1-P(G)_{\text{f}}$ (the QC-weight),which means that data which are almost certainly wrong ($P(G)_{\text{f}} \approx 1$) are given near-zero weight in the analysis. Data with a $P(G)_{\text{f}} > 0.75$ are considered 'rejected' and are flagged accordingly, for the purpose of diagnostics and feedback statistics, etc.

The normal definition of a cost function is

$$J_{\text{o}} = -\ln p \tag{5.6}$$

where p is the probability density function. Instead of the normal assumption of Gaussian statistics, we assume that the error distribution can be modelled as a sum of two parts: one Gaussian, representing correct data and one flat distribution, representing data with gross errors. We write

$$p_i = N_i[1-P(G_i)] + F_i P(G_i) \tag{5.7}$$

where subscript i refers to observation number i. N and F are the Gaussian and the flat distributions, respectively, given by

$$N_i = \frac{1}{\sqrt{2\pi}\sigma_{\text{o}}} \exp\left[-\frac{1}{2}\left(\frac{y_i - Hx}{\sigma_{\text{o}}}\right)^2\right] \tag{5.8}$$

$$F_i = \frac{1}{L_i} = \frac{1}{2l_i\sigma_{\text{o}}} \tag{5.9}$$

The flat distribution is defined over an interval L_i which in (5.9) has been written as a multiple of the observation error standard deviation σ_o. Substituting (5.7) to (5.9) into (5.6), we obtain after rearranging the terms, an expression for the QC-modified cost function J_{o}^{QC} and its gradient $\nabla J_{\text{o}}^{\text{QC}}$, in terms of the

normal cost function J_o^N

$$J_o^N = \frac{1}{2}\left(\frac{y_i - Hx}{\sigma_o}\right)^2 \tag{5.10}$$

$$J_o^{QC} = -\ln\left(\frac{\gamma_i + \exp[-J_o^N]}{\gamma_i + 1}\right) \tag{5.11}$$

$$\nabla J_o^{QC} = \nabla J_o^N \left(1 - \frac{\gamma_i}{\gamma_i + \exp[-J_o^N]}\right) \tag{5.12}$$

$$\nabla J_o^{QC} = \nabla J_o^N w_i \tag{5.13}$$

where

$$w_i = 1 - \frac{\gamma_i}{\gamma_i + \exp[-J_o^N]} \tag{5.14}$$

$$\gamma_i = \frac{P(G_i)/(2l_i)}{[1 - P(G_i)]/\sqrt{2\pi}} \tag{5.15}$$

5.3.2 Implementation

The *a priori* information, i.e. $P(G)_i$ and l_i, is set during the screening, in the routine DEPART, and stored in the NCMFGC1 and NCMFGC2-words of the ODB. Default values are set in DEFRUN, and can be modified by the namelist namjo. VarQC can be switched on/off for each observation type and variable individually using LVARQC, or it can be switched off all together by setting the global switch LVARQCG = .FALSE.. Since an as good as possible 'preliminary analysis' is needed before VarQC starts, it is necessary to perform part of the minimization without VarQC, and then switch it on.

HJO computes J_o^{QC} according to (5.11) and the QC-weight, w_i, according to (5.14).

The 3D-Var assimilation system (Chapter 3) can handle non-quadratic cost functions because it is using the M1QN3 minimization routine. This allows VarQC to be activated after NITERQC (40 by default) iterations during the minimization process. The 4D-Var assimilation system (Chapter 2) by default use a conjugate gradient minimization method that requires a strictly quadratic cost function. So in 4D-Var we use a quadratic formulation of VarQC, controlled by LQVARQC (default .TRUE.).

In the non-quadratic formulation the QC-weight, w_i, (5.14) is updated for each simulation based on the normalized departure values calculated from the latest model state. However, to obtain a strictly quadratic cost function the QC-weight is not allowed to change during the minimisation process, because the cost function shape would then change. Also, when LQVARQC = .TRUE. it is not possible to activate VarQC in the middle of the minimisation process, i.e. NITERQC MUST be zero, because this would otherwise introduce a sudden jump in the cost function. Therefore, in 4D-Var assimilations VarQC is not applied during the first outer-loop but only for the second (and possibly subsequent) outer-loop iteration(s).

The quadratic VarQC implementation satisfies the quadratic constraint by calculating the QC-weight, w_i, based on the high resolution trajectory fields and using this constant weight during the next minimisation step. The cost function value is increased/reduced for each simulation during the minimisation by a factor w_i^*(norm_dep_LR-norm_dep_HR), linearized in the vicinity of the high resolution cost function at norm_dep_HR. For the minimisation the most important input is the modification of the gradient by the weight w_i. During the trajectory run the high resolution departure is stored in ROBODY(..,MDBIFC1(NUPTRA+1)) in HDEPART. For each simulation during the next minimisation step the high resolution departure is read and normalized by the final observation error in HJO HJO loop 1.3.2. This is used as input to the VarQC weight calculations (loop 1.5.3–1.5.4 in HJO). It is assured that the same weight is used for each simulation in the minimisation if LLQVARQC_MIN = .TRUE., because ZFJOS_HR(JOBS,JBODY) is constant during the minimisation. The cost function value is calculated like Jo_varqc_LR = Jo_varqc_HR + w_i^*(norm_dep_LR-norm_dep_HR).

The *a posteriori* probability of gross error is stored in the ODB and passed to BUFR feedback. Storing in ODB is done at the final simulation of the last minimisation if LQVARQC = .F., but done during the final trajectory run if LQVARQC = .TRUE. in order to use the updated final trajectory for calculating the QC-weight based on final high-resolution analysis values. This is done in HJO – see Section 5.2.5.

Variational quality control for time correlated observations: The same method as above is (by default) applied for time correlated observations (LTC = .TRUE.). Here the high resolution departure values are required for all the time correlated observations. This is achieved by storing the normalized high resolution departures in the array RTCNDPHR in HJO and reading them in COMTC. They are then copied to local array ZTCNDPHR_W and then used instead of the low resolution departures in the calculation (overwriting ZOM1DEP_W by high resolution values).

5.4 OBSERVATION OPERATORS – GENERAL

The operator H is subdivided into a sequence of operators, each one of which performs part of the transformation from control variable to observed quantity.

(i) The inverse change of variable (CHAVARIN) converts from control variables to model variables.
(ii) The inverse spectral transforms put the model variables on the model's reduced Gaussian grid (controlled by SCAN2MDM).
(iii) A 12-point bi-cubic or 4-point bi-linear horizontal interpolation, similar to the semi-lagrangian interpolation routines described elsewhere, gives vertical profiles of model variables at observation points (controlled by COBS, COBSLAG). The surface fields are interpolated bi-linearly to avoid spurious maxima and minima. The three steps (i) to (iii) are common to all data types. Thereafter follows steps (iv) and (v).
(iv) Vertical integration of, for example, the hydrostatic equation to form geopotential, and of the radiative transfer equation to form radiances (if applicable).
(v) Vertical interpolation to the level of the observations.

The vertical operations depend on the variable. The vertical interpolation is linear in pressure for temperature (PPT) and specific humidity (PPQ), and it is linear in the logarithm of pressure for wind (PPUV). The vertical interpolation of geopotential (PPGEOP) is similar to wind (in order to preserve geostrophy) and is performed in terms of departures from the ICAO standard atmosphere for increased accuracy (Simmons and Chen, 1991, see Section 5.5 below). The current geopotential vertical interpolation together with the temperature vertical interpolation are not exactly consistent with hydrostatism. A new consistent and accurate vertical interpolation has been devised by Meteo-France, which may be important for intensive use of temperature information. The new routines have been tested by ECMWF and as the results were not unambiguously positive the new routines have not yet been adopted – and they are not described in this documentation. In the meantime, the old routines are still used (switch LOLDPP = .TRUE. in namct0), under the names PPT_OLD, PPGEOP_OLD and PPUV_OLD, with tangent linear PPTTL_OLD, PPGEOPTL_OLD and PPUVTL_OLD and adjoint PPTAD_OLD, PPGEOPAD_OLD and PPUVAD_OLD.

The vertical interpolation operators for SYNOP 10m wind (PPUV10M) and 2 m temperature (PPT2M) match an earlier version of the model's surface layer parametrisation. The vertical gradients of the model variables vary strongly in the lowest part of the boundary layer, where flow changes are induced on very short time and space scales, due to physical factors such as turbulence and terrain characteristics. The vertical interpolation operator for those data takes this into account following Monin–Obukhov similarity theory. Results using such operators, which follow Geleyn (1988) have been presented by Cardinali *et al.* (1994). It was found that 2 m temperature data could not be satisfactorily used in the absence of surface skin temperature as part of the control variable, as unrealistic analysis increments appeared in the near-surface temperature gradients. The Monin–Obukhov based observation operator for 10 m wind, on the other hand, is used for all 10m winds (SYNOP, DRIBU, TEMP, PILOT and SCAT).

Relative humidity is assumed constant in the lowest model layer to evaluate its 2m value (PPRH2M), see Subsection 5.8.4. The model equivalent of SSMI total column water vapour data is obtained by vertical integration of q (in GPPWC and PPPWC). The routine PPPWC is also used for vertical integration of GEMS trace gasses. Observation operators exist for precipitable water content (also using PPPWC) and thicknesses (PPGEOP).

The variational analysis procedure requires the gradient of the objective function with respect to the control variable. This computation makes use of the adjoint of the individual tangent linear operators,

applied in the reverse order. The details regarding observation operators for conventional data can be found in Vasiljevic *et al.* (1992), Courtier *et al.* (1998), and in the following sections.

5.5 THE OBSERVATION OPERATOR FOR CONVENTIONAL DATA

5.5.1 Geopotential height

The geopotential at a given pressure p is computed by integrating the hydrostatic equation analytically using the ICAO temperature profile and vertically interpolating $\Delta\phi$, the difference between the model level geopotential and the ICAO geopotential (Simmons and Chen, 1991). The ICAO temperature profile is defined as

$$T_{\mathrm{ICAO}} = T_0 - \frac{\Lambda}{g}\phi_{\mathrm{ICAO}} \tag{5.16}$$

where T_0 is 288 K, ϕ_{ICAO} is the geopotential above 1013.25 hPa and Λ is 0.0065 K m^{-1} in the ICAO troposphere and 0 in the ICAO stratosphere (the routine PPSTA). The ICAO tropopause is defined by the level where the ICAO temperature has reached 216.5 K (SUSTA). Using this temperature profile and integrating the hydrostatic equation provides T_{ICAO} and the geopotential ϕ_{ICAO} as a function of pressure (PPSTA). We may then evaluate the geopotential $\phi(p)$ at any pressure p following

$$\phi(p) - \phi_{\mathrm{surf}} = \phi_{\mathrm{ICAO}}(p) - \phi_{\mathrm{ICAO}}(p_{\mathrm{surf}}) + \Delta\phi \tag{5.17}$$

where p_{surf} is the model surface pressure and ϕ_{surf}, the model orography. $\Delta\phi$ is obtained by vertical interpolation from the full model level values $\Delta\phi_k$. The interpolation is linear in $\ln(p)$ up to the second model level (PPINTP) and quadratic in $\ln(p)$ for levels above it (PPITPQ, see below). Following Simmons and Burridge (1981) the full model level values are obtained by integrating the discretized hydrostatic equation using the routine GPGEO of the forecast model to give

$$\Delta\phi_k = \sum_{j=L}^{k+1} R_{\mathrm{dry}}(T_{\mathrm{v}_j} - T_{\mathrm{ICAO}_j}) \ln\left(\frac{p_{j+1/2}}{p_{j-1/2}}\right) + \alpha_k R_{\mathrm{dry}}(T_{\mathrm{v}_k} - T_{\mathrm{ICAO}_k}) \tag{5.18}$$

with

$$\alpha_k = 1 - \frac{p_{k-1/2}}{p_{k+1/2} - p_{k-1/2}} \ln\left(\frac{p_{k+1/2}}{p_{k-1/2}}\right)$$

for $k > 1$ and $\alpha_1 = \ln(2)$.

(a) Quadratic vertical interpolation near the top of the model

Above the second full level of the model, the linear interpolation (PPINTP) is replaced by a quadratic interpolation in $\ln p$, performed in the routine PPITPQ using

$$z(\ln p) = a + b(\ln p) + c(\ln p)^2 \tag{5.19}$$

where a, b and c are constants determined so that the above equation fits the heights at the top levels ($k = 1, 2$ and 3). The interpolation formula is

$$\phi(\ln p) = z_2 + \frac{(z_2 - z_1)(\ln p - \ln p_2)(\ln p - \ln p_3)}{(\ln p_2 - \ln p_1)(\ln p_1 - \ln p_3)} - \frac{(z_2 - z_3)(\ln p - \ln p_1)(\ln p - \ln p_2)}{(\ln p_2 - \ln p_3)(\ln p_1 - \ln p_3)} \tag{5.20}$$

where 1,2 and 3 refer to levels $k = 1, 2$ and 3, respectively.

(b) Below the model's orography

The extrapolation of the geopotential below the model's orography is carried out as follows: Find T^* (surface temperature) by assuming a constant lapse rate Λ, from the model level above the lowest model level (subscript $l-1$), see the routine CTSTAR, using

$$T^* = T_{l-1} + \Lambda\frac{R_{\mathrm{dry}}}{g}T_{l-1}\ln\frac{p_{\mathrm{surf}}}{p_{l-1}} \tag{5.21}$$

$$T^* = \frac{\{T^* + \max[T_y, \min(T_x, T^*)]\}}{2} \tag{5.22}$$

Find the temperature at mean sea level, T_0 (also in CTSTAR) from

$$T_0 = T^* + \Lambda \frac{\phi_{\text{surf}}}{g} \tag{5.23}$$

$$T_0 = \min[T_0, \max(T_x, T^*)] \tag{5.24}$$

where T_x is 290.5 K and T_y is 255 K. The geopotential under the model's orography is (in PPGEOP) calculated as

$$\phi = \phi_{\text{surf}} - \frac{R_{\text{dry}} T^*}{\gamma}\left[\left(\frac{p}{p_{\text{surf}}}\right)^{\gamma} - 1\right] \tag{5.25}$$

where $\gamma = \frac{R_{\text{dry}}}{\phi_{\text{surf}}}(T_0 - T_{\text{surf}})$.

5.5.2 Wind

In PPUV a linear interpolation in $\ln p$ (PPINTP) is used to interpolate u and v to the observed pressure levels up to the second full model level, above which a quadratic interpolation is used (PPITPQ, see Subsection 5.5.1). Below the lowest model level wind components are assumed to be constant and equal to the values of the lowest model level.

5.5.3 Humidity

Specific humidity q, relative humidity U and precipitable water content PWC are linearly interpolated in p, in PPQ, PPRH and PPPWC, respectively. Upper air relative humidity data are normally not used, but could be used, if required. The use of surface relative humidity data is described in Subsection 5.8.4.

(a) Saturation vapour pressure

The saturation vapour pressure $e_{\text{sat}}(T)$ is calculated using Tetens's formula given by

$$e_{\text{sat}}(T) = a_1 \exp^{a_3\left(\frac{T-T_3}{T-a_4}\right)} \tag{5.26}$$

using FOEEWM (mixed phases, water and ice) in the model and FOEEWMO (water only) for observations. The use of water-phase only is in accordance with the WMO rules for radiosonde and synop reporting practices. Note that these statement functions compute $(R_{\text{dry}}/R_{\text{vap}})e_{\text{sat}}(T)$, with the parameters set according to Buck (1981) and the AERKi formula of Alduchov and Eskridge (1996), i.e. $a_1 = 611.21$ hPa, $a_3 = 17.502$ and $a_4 = 32.19$ K over water, and for FOEEWM $a_3 = 22.587$ and $a_4 = -0.7$ K over ice, with $T_3 = 273.16$ K. Furthermore in FOEEWM the saturation value over water is taken for temperatures above 0°C and the value over ice is taken for temperatures below −23°C. For intermediate temperatures the saturation vapour pressure is computed as a combination of the values over water $e_{\text{sat(water)}}$ and $e_{sat(ice)}$ according to the formula

$$e_{\text{sat}}(T) = e_{\text{sat(ice)}}(T) + [e_{\text{sat(water)}}(T) - e_{\text{sat(ice)}}(T)]\left(\frac{T - T_i}{T_3 - T_i}\right)^2 \tag{5.27}$$

with $T_3 - T_{\text{i}} = 23$ K.

(b) Relative humidity

In GPRH relative humidity U is computed from

$$U = \frac{pq\frac{R_{\text{vap}}}{R_{\text{dry}}}}{\left[1 + \left(\frac{R_{\text{vap}}}{R_{\text{dry}}} - 1\right)q\right]e_{\text{sat}}(T)} \tag{5.28}$$

and then in PPRH interpolated to the required observed pressure levels (using PPINTP). Below the lowest model level and above the top of the model is U assumed to be constant. Saturation vapour pressure is calculated using FOEEWMO if GPRH has been called form the observation operator routines, and using FOEEWM if called from the model post processing.

(*c*) *Precipitable water*

In GPPWC precipitable water is calculated as a vertical summation from the top of the model by

$$PWC_k = \frac{1}{g}\sum_{i=1}^{k} q_i(p_i - p_{i-1}) \tag{5.29}$$

and then in PPPWC interpolated to the required observed pressure levels (using PPINTP). PWC is assumed to be zero above the top of the model. Below the model's orography PWC is extrapolated assuming a constant $q = q_l$.

(*d*) *Specific humidity*

Specific humidity q is in PPQ interpolated to the required observed pressure levels (using PPINTP). Below the lowest model level and above the top of the model is q assumed to be constant and equal to q_l and q_1, respectively.

5.5.4 Temperature

Temperature is interpolated linearly in pressure (PPINTP), in the routine PPT. Above the highest model level the temperature is kept constant and equal to the value of the highest model level. Between the lowest model level and the model's surface the temperature is interpolated linearly, using

$$T = \frac{(p_{\rm surf} - p)T_l + (p - p_l)T^*}{p_{\rm surf} - p_l} \tag{5.30}$$

Below the lowest model level the temperature is extrapolated by

$$T = T^*\left[1 + \alpha \ln\frac{p}{p_{\rm surf}} + \frac{1}{2}\left(\alpha \ln\frac{p}{p_{\rm surf}}\right)^2 + \frac{1}{6}\left(\alpha \ln\frac{p}{p_{\rm surf}}\right)^3\right] \tag{5.31}$$

with $\alpha = \Lambda R_{\rm dry}/g$, for $\phi_{\rm sat}/g < 2000$ m, but α is modified for high orography to $\alpha = R_{\rm dry}(T_0' - T^*)/\phi_{\rm surf}$, where

$$T_0' = \min(T_0, 298) \tag{5.32}$$

for $\phi_{\rm surf}/g > 2500$ m, and

$$T_0' = 0.002[(2500 - \phi_{\rm surf}/g)T_0 + (\phi_{\rm surf}/g - 2000)\min(T_0, 298)] \tag{5.33}$$

for $2000 < \phi_{\rm surf}/g < 2500$ m. If $T_0' < T^*$ then α is reset to zero. The two temperatures T^* and T_0 are computed using (5.21) to (5.24).

5.6 SATELLITE RADIANCE OPERATORS

The majority of satellite data assimilated currently is radiances. Radiances, rather than retrieved products, are assimilated directly (Andersson *et al.*, 1994), wherever possible. The current operational configuration uses clear level-1C radiances from a number of sensors, including AIRS, SSMI, ATOVS (McNally *et al.*, 1999), as well as geostationary water vapour clear-sky radiances (Munro *et al.*, 2004). Rain affected SSMI radiances are also assimilated (Bauer *et al.*, 2006a).

There are currently two routes in the IFS to assimilate satellite radiances. The first one uses the radiative transfer operator on model profiles interpolated to observation locations. The processing within 3D/4D-Var of satellite data assimilated through this route follows the same general layout as that of conventional data. This route is currently used for all clear-sky radiances. The second route uses the radiative transfer operator on model profiles at model grid point locations, and associates the output with observations within a given collocation radius around the model grid point. This assimilation route differs considerably from that used for conventional observations. The route is currently used for rain- or cloud-affected microwave radiances only.

For the naidr radiances, the observation operators for both routes are different flavours of the RTTOV radiative transfer model (Saunders and Matricardi, 1998; Matricardi *et al.*, 2001), currently using RTTOV version 8.

5.6.1 Common aspects for the setup of nadir radiance assimilation

The operational radiance assimilation shares the following setup aspects for both radiance assimilation routes. The datasets are distinguished by a satellite ID, a sensor ID, and a codetype. The latter is used to distinguish clear-sky (codetype=210=NGTHRB) or rain-affected radiances (codetype=215=NSSMI).

The main set-up routine for radiances is SURAD. It recognises satellite IDs (call to GETSATID), reads RTTOV coefficient files (call to RTSETUP), and builds a "satellite group table" containing information on which satellite groups are present.

(*a*) *Satellite identifiers and sensors*

Satellite identifiers are dealt with in the routine GETSATID, called from SURAD. The ODB contains the identifiers as given in the original BUFR messages. Lists of identifiers for which data exist in any given ODB are prepared in the routine SURAD. The routine GETSATID matches those BUFR satellite identifiers with the more traditional satellite numbers, used by the RT-code (e.g. 10 for NOAA-10 and 5 for METEOSAT-5). The id-conversion tables can be modified through the namelist NAMSATS.

The various types of radiance data in the IFS are also classified by sensor. Each satellite sensor is assigned a number, defined in the module YOMTVRAD. The sensor number is used as index to various tables containing observation errors, BgQC thresholds, VarQC parameters, the J_o-table JOT, etc. See the routine DEFRUN.

(*b*) *Satellite group table*

Various satellite-related indices are gathered in the routine SURAD in the FORTRAN90 data structure called the 'satellite group table', satgrp_t (defined in YOMTVRAD). The table contains elements such as the satellite ID, the sensor ID, the codetype, a sequence number for addressing the transmittance coefficients (rtcoef_pos), the number of channels, a channel number list, etc. The various satellite-related indices are universally determined across all processors. There is one entry in the satellite group table per satellite, sensor, and codetype. A list of all the satellite groups can be found in the *ifstraj* output by searching for SATGRP.

(*c*) *Fixed pressure levels and RT validation bounds*

Aspects of the clear-sky RTTOV calculations are set-up in the call to RTSETUP from SURAD. The routine RTSETUP calls RTTVI (in the satrad library) which reads in the transmittance coefficients to memory for the satellites present. There is one file containing these coefficients for each instrument and satellite. The files can be found under */home/rd/rdx/data/31r1/sat/rttov8*.

The list of the NLSAT = 43 fixed pressure levels used in clear-sky RTTOV calculations is passed from the RTTOV library satrad (where they have been read from the transmittance coefficients file) to RTSETUP and SURAD and copied to YOMTVRAD. RTSETUP also similarly obtains (from RTTOV) lists of temperature, humidities and ozone limits indicating the valid range of the RT transmittance regression. The RT code is not reliable in conditions beyond these limits. Checks are applied in RADTR. The IFS uses just one set of fixed pressure levels, reference or limit profiles (rather than having instrument dependent ones), and the ones that are read last are used.

5.6.2 Clear-sky nadir radiances

(*a*) *Radiance observation errors, bias and emissivity*

Observation errors for 1C radiances are written to the ODB in a call to RAD1COBE (from HRETR). The bias correction is calculated by variational methods (see below) and stored in the ODB. Microwave (EMIS_MW_N) and infrared (EMIS_IR) emissivity are computed in RAD1CEMIS and stored in the ODB for later use by RTTOV. The routine EMIS_IR uses the regression of ISEM-6 (Sherlock, 1999) to compute emissivities for the infrared channels of VTPR, HIRS, Meteosat, GOES, and AIRS.

(*b*) *Calling the radiative transfer model*

The routine HOP interpolates the model profiles of temperature, humidity and ozone (T, q, and oz) to the NLSAT(= 43) RT levels ($\dot{T}$ and $\dot{q}$) and calls the interface RADTR to the RT code RTTOV. The standard routines PPT and PPQ are used to carry out the vertical interpolation, and they are called through the PPOBSA interface, as usual. Various radiance preparations have been gathered in the routine HRADP. In HRADP, the model's pressure at the surface height of the observation location (given in the report) is calculated, using PPPMER.

For the purpose of radiance calculations $T_{2\mathrm{m}} = T_l$ (the lowest model level temperature) and $q_{2\mathrm{m}} = \dot{q}_{NLSAT}$ (specific humidity at the lowermost of the RT pressure levels). These quantities represent a very shallow layer of air near the surface and contribute little to the calculated radiances – it was not considered necessary to use PPT2M and PPRH2M (Section 5.8) in this context. In order to make the radiance cost function continuous in p_{surf} it was necessary to ensure that $\dot{T}$ and $\dot{q}$ approach $T_{2\mathrm{m}}$ and $q_{2\mathrm{m}}$ as the pressure on any of the RT levels approaches p_{surf}. This is done in a section of HRADP. More details on the radiative transfer code RTTOV can be found in Eyre (1991), updated by Saunders and Matricardi (1998).

In HOP the observation array is searched for radiance data. The CCMA ODB (after screening) contains only those data to be used by the analysis. A list of existing channel numbers for each report is constructed. Model radiances for exactly those channels are then requested from the RT-code, via the interface RADTR. The routine RADTR checks that the input model profile is within the valid range of the transmittance regression. It packets the profiles into chunks of work of the appropriate maximum size for the RT-code (currently 65). The RT packet size has been communicated to IFS in the call to RTSETUP. The output is radiances for the channels requested.

The tangent linear HOPTL and the adjoint HOPAD follow the same pattern as HOP. In both the TL and the adjoint $\dot{T}$ and $\dot{q}$ have to be recomputed before the actual tangent linear and adjoint computations can start. The pointers to the radiance data in observation array are obtained just as it was done in the direct code.

(*c*) *Skin-temperature 'sink-variable' at satellite FOVs*

In the case of 1C, or 'raw' radiance data, as used since May 1999 (McNally *et al.*, 1999) surface skin temperature is retrieved by 3D/4D-Var at each field of view, if the switch LTOVSCV is on (default is on), in namelist NAMVAR. This is done for all infrared and microwave satellite sensors and instruments. The handling of the SST retrieval at the radiance FOVs is performed in the routine HRADP/HRADPTL/HRADPAD, called from HOP/HOPTL/HOPAD. The background SST is provided by the model trajectory integration, and a background error of 1K/5K is assigned over sea/land, respectively (stored in TOVSCVER, module YOMTVRAD). The gradient with respect to SST obtained from RTTOV is temporarily stored in the TOVSCVX array and later transferred to its location in the distributed control vector (Chapter 2). The next iteration of the minimisation provides updated SST increments (also stored in TOVSCVX) that are used by RTTOVTL in subsequent iterations. The outer-loop iterations result in a new linearisation state SST, stored in TOVSCVX5. All the SST-related information at FOV locations that needs to be passed between job-steps, reside in the ODB, in the satellite predictors table. This approach has been adopted for CO2 retrieval at AIRS FOVs (Engelen *et al.*, 2004).

(*d*) *Variational bias correction*

For clear-sky radiances (codetype 210), variational bias correction (VarBC, Dee (2004)) is used operationally since 31r1 (switch LVARBC in namelist NAMVAR). VarBC estimates the coefficients used in the radiance bias models within the main analysis by making these coefficients part of the control vector. A separate user's guide is available for VarBC (Auligné and Dee, 2006), and only the main aspects are summarised here.

VarBC performs the bias correction on the basis of VarBC groups, defined by satellite, sensor, codetype, and channel, and different bias models etc can be used per group. General VarBC options are handled through the namelist NAMVARBC. Information produced by VarBC (coefficients, histograms) is stored in the VARBC.cycle file which also facilitates the cycling of coefficients from one cycle to the next. The

file gets written and read in the routines WRVARBC and RDVARBC, respectively. The information from the VARBC.cycle file is held in memory in the YVARBC structure defined in the module YOMVARBC.

The configuration of VarBC is done in the routine SUVARBC where the bias models are defined by VarBC group. The routine initialises the YVARBC structure depending on what is available in the VARBC.cycle file. If a new VarBC group appears, a "cold start" is performed for this group in the routine CSVARBC, depending on the setting of NCS_CONFIG specified in the namelist NAMVARBC. Options are 1) setting all coefficients to zero (= 0), 2) initialise coefficients from the old-style Harris-and-Kelly files (= 1), or set the constant offset used by VarBC to the mode of the uncorrected First Guess departures (= 2).

The actual bias estimation is performed in the routine RADTRB and its tangent linear and adjoint. It is called from HOP/TL/AD and HRETR. It also calls the routine BIASPRED which calculates the predictors used by VarBC, and accumulates some statistics for these predictors. These statistics are used in the preconditioning for VarBC (see routine CVARBC). Cost function contributions from VarBC are calculated in FJVARBC.

5.6.3 Cloud/rain affected nadir radiances

The implementation of cloud and rain-affected microwave radiances (SSM/I) was first introduced in IFS model cycle 29R2 in June 2005 and has been revised with cycle 31R1 in September 2006. Rain-affected radiances are used within a 1D+4D-Var methodology that is described in detail in Bauer *et al.* (2006a) and Bauer *et al.* (2006b). The 1D-Var algorithm retrieves total column water vapour (TCWV) in clouds and precipitation using SSM/I radiance observations and with profiles of temperature, specific humidity and 10-metre wind speed in the control vector. The TCWV observations are then assimilated like other observations in 4D-Var.

The 1D+4D-Var method is activated (in analysis experiments) in prepIFS section 'Satellites' by setting LRAIN1D to .TRUE. (default). LRAIN1D further sets LERAIN that is used in namelist NAEPHY of the model physics package (see below), and LERAIN_ACTIVE in the master routine for the grid-point part of the model calculations (IFS/CONTROL/gp_model). A 1D-Var-specific namelist is contained in scripts *ifstraj* and is called NAMONEDVAR. The entries are declared and transferred to IFS by IFS/MODULES/yomonedvar and IFS/NAMELIST/namonedvar.

(*a*) *Pre-processing*

Since SSM/I observations are used in both clear skies and clouds/precipitation, a separate routine for the pre-processing of rain affected SSM/I radiances is required. In program SATRAD/PRESCREEN/bufr_screen_ssmi_1d the ssbt BUFR files are decoded and reduced to data over oceans (NESDIS surface flag equal to 5) and by deactivating the pre-thinning. The original SSM/I BUFR subtype 127 is used (different from SSM/I clear-sky radiances which are set to BUFR subtype 55 to become codetype 210) which associates this data with IFS codetype 215 (see ODB/CMA2ODB/buf2cmat_new) that is TCWV-observations.

In contrast to all other observations used in the IFS, observations are assimilated with reference to single model grid points rather than by interpolation from multiple points to observation location. This is because (1) the observation operator requires numerous two and three-dimensional model fields as input and (2) because either the interpolation of operator input fields to run the observation operator at observation location or the interpolation of operator output to observation location produce unacceptable errors. The 1D-Var is therefore run at model grid-points along the first model trajectory at full model resolution and for each time step.

The SSM/I observations are retrieved from ODB and associated with model grid-points that are located within a maximum search radius of 10 km (7 km) for wavenumber cut-offs less/equal 511 (above), respectively. Only one observation is associated with one grid-point per time step. This operation is performed in IFS/VAR/gp_ssmi before the call to IFS/PHYS_EC/ec_phys and by first calling IFS/VAR/gp_nearest and then by calling IFS/VAR/gp_ssmi_obs2gp and IFS/VAR/gp_ssmi_iobs2gp. IFS/VAR/gp_ssmi is called from the main grid-point model routine IFS/CONTROL/gp_model.

In the same section of IFS/CONTROL/gp_model the load-balancing for the 1D-Var is activated if LERAIN_LB=.TRUE. (default). This flag is also set in NAMONEDVAR. The load-balancing reduces the considerable computational cost of the 1D-Var retrievals by about 50% through the work-load dependent redistribution of 1D-Var retrievals across the MPI tasks. Routines IFS/ONEDVAR/onedvar_raintb_prc and IFS/ONEDVAR/onedvar_raintb_set set up the helper processes after establishing the work-load. The helper routine IFS/ONEDVAR/onedvar_raintb_hlp calls the main 1D-Var routine IFS/ONEDVAR/onedvar_raintb for each helper and IFS/ONEDVAR/onedvar_raintb_snd and IFS/ONEDVAR/onedvar_raintb_rcv send and receive the input and output of IFS/ONEDVAR/onedvar_raintb to and from the individual processors. After the call to IFS/PHYS_EC/ec_phys the output from the 1D-Var is sent back to the ODB at the appropriate observation location in IFS/VAR/gp_ssmi_inv ready for use in IFS/PP_OBS/hop.

Note that the 1D-Var does not perform a minimization of multiple profiles at once and is therefore not suited for vector machines. However, on multiple processor machines and with active load-balancing, the computational cost is acceptable given the complexity of the observation operator and the 1D-Var minimization.

(b) *1D-Var retrieval*

Set-up: As for other (satellite) observations, several coefficient and parameter files are required and are loaded before the model run. The set-up is performed in the pre-screening and by calling IFS/ONEDVAR/onedvar_setup from IDS/VAR/rtsetup. This is done only from the master MPI task to avoid excessive I/O.

Despite being satellite observations, the SSM/I radiances used in the 1D-Var retrieval do not enter the variational bias correction. For bias correction, a scan-bias correction is performed that is obtained from clear-sky observations over at least 3 months of data and from off-line calculations. The scan-bias correction is assumed to be constant per satellite and is stored per satellite, radiometer channel and scan position in */home/rd/rdx/data/sat/scanbias.ssmi*. This correction is directly applied to the BUFR SSM/I radiances in SATRAD/PRESCREEN/bufr_screen_ssmi_1d. In additiona, an air-mass dependent bias correction is performed (see below).

IFS/ONEDVAR/onedvar_setup performs the reading of the air-mass dependent bias correction coefficients that are also satellite and channel dependent and stored in */home/rd/rdx/data/sat/biascoef.ssmi*. Predictors are TCWV, sea-surface wind speed, and rain water path, i.e. there are four coefficients per channel. The air-mass bias correction accounts for the geographical variability of model background error standard deviations of temperature and specific humidity by calling IFS/ONEDVAR/onedvar_get_bgsig for each profile from IFS/ONEDVAR/onedvar_raintb. This routine interpolates the operational low-resolution (T95) horizontal and vertical short-range forecast errors to the high-resolution grid-point location with IFS/VAR/suhifce and IFS/VAR/suvifce.

The same routine also calls IFS/ONEDVAR/onedvar_get_bgcor that reads the correlation matrices for model background errors of temperature and specific humidity contained in */home/rd/rdx/data/sat/cor_t_*norm* and */home/rd/rdx/data/sat/cor_q_*norm*. The '*' refers to the number of levels (31, 60, 91). It is assumed that the vertical correlations are globally constant.

IFS/ONEDVAR/onedvar_setup also performs the reading of the clear-sky radiative transfer coefficients and the coefficients required for the scattering calculations in RTTOV that is carried out from IFS/ONEDVAR/onedvar_get_rtcoeff. These coefficients are contained in */home/rd/rdx/data/sat/rt_coef73/rtcoef_dmsp_*_ssmi.dat* (read by SATRAD/RTTOV/rttov_readcoeffs) and */home/rd/rdx/data/sat/mietable_dmsp_ssmi* (read by SATRAD/RTTOV/rttov_readscattcoeffs) of which the former are dependent on the DMSP satellite ('*' being between 8 and 15).

The same set-up routine also reads the namelist NAMONEDVAR after having initialized the namelist entries with default values.

Screening: IFS/ONEDVAR/onedvar_raintb is the main driver routine for the 1D-Var and calls IFS/ONEDVAR/onedvar_screen to determine for which profiles to perform the retrieval. IFS/ONEDVAR/onedvar_screen excludes observations that are assumed erroneous (land/sea-ice contaminated, radiances out of bounds etc.) and performs two threshold checks to detect the presence of clouds and/or precipitation. These threshold checks are identical to those applied in IFS/PP_OBS/rad1cemis to ensure that each SSM/I observation is only used once, either in the clear-sky radiance assimilation or the 1D+4D-Var framework.

The air-mass bias-correction is also performed in IFS/ONEDVAR/onedvar_screen by IFS/ONEDVAR/onedvar_get_bias through a linear regression-based correction using the above mentioned predictors and the coefficients obtained from */home/rd/rdx/data/sat/biascoef.ssmi*.

Minimizer: The 1D-Var employs the M1QN3 minimizer in the retrieval that is a quasi-Newton iterative method (IFSAUX/MINIM/m1qn3_1dv and associated '*1dv' routines in the same directory (Gilbert and Lemaréchal, 1989). The inversion problem is pre-conditioned by a change of variable because of the large differences between temperature and specific humidity in terms of dynamic range and magnitude. Prior to the minimization, an initial call to the observation operator interface routine IFS/ONEDVAR/onedvare_simul is carried out to identify the direction of the largest gradient in cost-function-space.

IFS/ONEDVAR/onedvar_simul performs the change of variable and calls both forward and adjoint observation operator, namely IFS/ONEDVAR/onedvar_obsop and IFS/ONEDVAR/onedvar_obsop_grad. ¿From these, the updated state vector and the cost-function elements are called per simulation (gradient search) and per iteration.

Observation operator: The observation operator is organized such that it has a component in both the IFS and SATRAD parts of the code. This was necessary because of the uni-directional dependencies in the IFS-code that, e.g., do not allow SATRAD-code modules to be visible from IFS-code elements. The forward, tangent-linear (not active) and adjoint versions of the observation operator are contained in IFS/ONEDVAR/onedvar_obsop, IFS/ONEDVAR/onedvar_obsop_tl and IFS/ONEDVAR/onedvar_obsop_grad. IFS/ONEDVAR/onedvar_obsop calls the linearized moist physics parameterizations for convection IFS/PHYS_EC/cucalln2 (Lopez and Moreau, 2005) and large-scale condensation IFS/PHYS_EC/cloudst (Tompkins and Janisková, 2004) via the interface-routine IFS/PHYS_EC/phys_nl. The tangent-linear and adjoint versions are called from the corresponding routines IFS/PHYS_EC/phys_tl and IFS/PHYS_EC/phys_ad.

These routines calculate profiles of cloud cover, cloud liquid water and ice as well as liquid and frozen precipitation fluxes from profiles of temperature, specific humidity as well as their tendencies, surface latent and sensible heat-flux, surface friction velocity, and pressure profile. IFS/ONEDVAR/onedvar_obsop then calls SATRAD/ONEDVAR/onedvar_obsop_rttov for performing the radiative transfer calculations and to feed back the model radiances to IFS/ONEDVAR/onedvar_simul or IFS/ONEDVAR/onedvar_raintb.

IFS/ONEDVAR/onedvar_obsop_grad first reapplies the forward physics model, then calls SATRAD/ONEDVAR/onedvar_obsop_grad_rttov before calling IFS/PHYS_EC/phys_ad to obtain the cost-function gradients that correspond to SSM/I microwave radiance departures.

SATRAD/ONEDVAR/onedvar_obsop_rttov represents the interface-routine of the 1D-Var with the multiple-scattering radiative transfer code that is part of RTTOV (for details see Bauer *et al.* (2006c)). This requires the initialization of arrays that configure the active radiometer channels per profile and instrument (SATRAD/ONEDVAR/ondevar_set_rtindex). The clear-sky radiative transfer model input arrays (temperature, specific humidity) are interpolated to the RTTOV reference levels by SATRAD/RTTOV/rttov_intex prior to the call of the main scattering model routine SATRAD/RTTOV/rttov_scatt. For the adjoint calculations, SATRAD/ONEDVAR/onedvar_obsop_grad_rttov first duplicates the forward calculations as in SATRAD/ONEDVAR/onedvar_obsop_rttov before calling SATRAD/RTTOV/rttov_scatt_ad and SATRAD/RTTOV/rttov_intex_ad, respectively.

Note that the scattering calculations are performed at model levels while the clear-sky calculations are performed for RTTOV-levels and have to be interpolated before and after the call of RTTOV. For increased efficiency, SATRAD/RTTOV/rttov_scatt shares two model components with the clear-sky part of the code that is the computation of clear-sky atmospheric transmittances (SATRAD/RTTOV/rttov_direct) and the computation of sea-surface emissivity as a function of microwave channel, 10-metre wind speed and sea-surface temperature (SATRAD/RTTOV/rttov_calcemis_mw).

(*c*) *Post-processing*

Once the 1D-Var minimization has been successfully completed, several quality control checks are performed and failure codes are defined for further diagnostic analyses. IFS/ONEDVAR/onedvar_obsop is called another time to compute the analysis SSM/I radiances that are further used in a convergence check. The output TCWV is calculated as well as the TCWV 1D-Var analysis error assuming a linear system that is from $A^{-1} = B^{-1} + H^T R^{-1} H$ with background moisture error covariance matrix, $\mathbf{B}$, observation error covariance matrix, $\mathbf{R}$, and Jacobian matrix with respect to moisture, $\mathbf{H}$.

Optionally (driven by namelist entry LERAIN_LINTEST from NAMONEDVAR), a linearity test of the entire observation operator can be performed by a call to IFS/ONEDVAR/onedvar_lintest that calls IFS/ONEDVAR/onedvar_obsop and IFS/ONEDVAR/onedvar_obsop_tl to evaluate tangent-linear against finite-difference forward model output.

To store the 1D-Var output in ODB for further use in the 4D-Var system, IFS/CONTROL/gp_model calls IFS/VAR/gp_ssmi_inv after the execution of IFS/PHYS_EC/ec_phys. This process is finished once the first model trajectory has been calculated and before any other data type has been touched. The failure flags are stored in ODB (see information in IFS/ONEDVAR/onedvar_raintb).

(*d*) *4D-Var assimilation*

The 4D-Var system treats the 1D-Var TCWV retrievals like all other observations. This means they are handled as single-level observations in the main screening and operator routines IFS/PP_OBS/hretr and IFS/PP_OBS/hop (also TL and AD). Since data screening has already been performed inside the 1D-Var, IFS/PP_OBS/hretr only stores model top and bottom pressure in ODB for the vertical moisture integration.

In IFS/PP_OBS/hop, it is verified that observations with codetype=215 (i.e. IOBSTYPE=NSSMI) are associated with TCWV (IVNMRQ(JOBS,JBODY) = NVNUMB(6)) and not with radiances (IVNMRQ(JOBS,JBODY) == NVNUMB(50)) and that the model fields are interpolated like other single-level observations. The only specific operation with respect to TCWV in the observation operator routines is made in IFS/PP_OBS/hdepart where the first-guess departures in the screening are calculated using first-guess TCWV values that were produced in the 1D-Var code and stored in ODB.

5.6.4 Clear-sky limb radiances

Assimilation of clear-sky limb radiances has been implemented in the IFS for experimental purposes. The radiances are assimilated using the RTLIMB radiative transfer model which is an extention of RTTOV to the limb geometry. Details of the radiative transfer model and the assimilation of limb radiances can be found in Bormann *et al.* (2005); Bormann and Healy (2006); Bormann and Thépaut (2007); Bormann *et al.* (2007). Many aspects have been primarily developed for the assimilation of MIPAS limb radiances; the assimilation of radiances from other sensors is likely to require additional coding.

Limb radiances fall under obstype 10 “Limb observations”, codetype 251. The general approach mirrors that used for clear-sky nadir radiances, i.e., the assimilation uses spatially interpolated vertical profile(s) of model variables. However, setup routines, the radiative transfer code, and the interface routines are different from the clear-sky nadir radiance assimilation.

The routine SULIMB sets up a limb group table (defined and stored in module YOMLIMB), for each satellite, sensor, and codetype, along the lines of the satellite group table set up in SURAD. Limb radiances are treated together with GPS radio occultation observations here. SULIMB calls the routine

RTL_SETUP which includes the interface routine to the satrad library to read the RTLIMB coefficient files. Note that in contrast to the setup for nadir radiances, the code is set up to use fixed pressure levels, reference and limit profiles that are specific to the RT-coefficient files. This information is stored in the limb group table. RTL_SETUP also reads a channel selection file into the channel selection structure Y_LIMB_CHAN_SEL in the module YOMLIMB. Observation errors and constant, channel-specific biases are also read here from auxillary files and stored in dedicated variables YOMLIMB.

Setting of observation errors and biases and screening of clear-sky limb radiances is performed from HRETR in the routines RTL_OBERROR and RTL_SCREEN. The latter applies the channel selection previously stored in Y_LIMB_CHAN_SEL in module YOMLIMB, and it performs cloud screening.

The actual assimilation happens in the routine HOP and its tangent linear and adjoint. HOP calls the routine RTL_HOP_1D which performs the following tasks: it interpolates the model profiles in the vertical to the fixed pressure levels (using the standard interpolation routines), does a simple extrapolation above the model top if required (based on a fixed mesospheric lapse rate for temperature, and holding humidity or ozone constant), and it checks the model profiles against the validity limits provided with the RTLIMB coefficient file. Finally, RTL_HOP_1D calls RTLIMB_HAT to enter the satrad library and perform the radiance computations.

The routine RTL_HOP_1D is used when local horizontal homogeneity is to be assumed for the radiative transfer computations. Alternatively, the radiative transfer computations can take the horizontal structure in the limb-viewing plane into account by providing a series of profiles covering the limb-viewing plane. In this case, profiles provided by PREINT2D are used, and the routine RTL_HOP_2D is called from HOP instead of RTL_HOP_1D. The 2d facility is switched on by specifying NOBSPROFS(10) $>$ 1 for obstype 10 in the namelist NAMNPROF. Note that this means GPS radio occulation bending angles present in the assimilation will also take horizontal structure into account.

5.7 OTHER SATELLITE OBSERVATION OPERATORS

Non-radiance satellite data are assimilated through various other observation operators, using model profiles interpolated to observation locations. Different operators are called from HOP, depending on the variable identifier of the observation. Dedicated observation operators exist, for instance, for scatterometer ambiguous surface winds (Stoffelen and Anderson, 1997; Isaksen and Janssen, 2004), or SATEM thicknesses and PWC (Kelly and Pailleux, 1988; Kelly *et al.*, 1991), SSM/I total-column water vapour (TCWV) and wind speed.

5.7.1 Atmospheric Motion Vectors

Groups of AMVs (aka SATOBs) are set up in the routine SUAMV, one group per satellite, computational method, and codetype. The information is stored in the satob group table, residing in the module YOMTVRAD.

The group table also specifies what type of observation operator is to be used for the particular group (entry obs_oper). The default used in operations is to assimilate all AMVs as single-level wind observations (Tomassini *et al.*, 1997; Bormann *et al.*, 2003), much in the same way as conventional data, using the same interpolation routine. Other options are to treat AMVs as layer averages, with weights specified by Gaussian functions around the assigned or otherwise specified reference level, or by using weighting functions obtained from radiative transfer calculations. The latter are made available in the routine HRETR.

5.7.2 Thicknesses

The pressures of layer bounds (top T, and bottom B) are found (in HOP) by scanning the observation array for thickness data. The geopotential for the top and the bottom of the layer are computed, using PPGEOP (Section 5.5), and the thickness is given by the difference $\phi_T - \phi_B$.

5.7.3 Precipitable water from SATEM and SSM/I

As for thicknesses, the pressures of layer bounds are found by scanning the observation array for TOVS PWC data. For SSMI TPW, the top pressure is set to the top of the model and the lower pressure bound is p_s. The PWC for the top and the bottom of the layer are computed, using PPPWC (Subsection 5.5.3), and the layer PWC is given by the difference $\mathrm{PWC_B} - \mathrm{PWC_T}$.

5.7.4 Ozone-layer retrievals from satellites

Ozone is used in the form of integrated layers bounded by a top and bottom pressure which are given as a part of the observation. The unit for ozone is kgm^{-2}. For ozone the same observation operator is used as for precipitable water (PPPWC, Section 5.4). The same concept is applied to all data, whethertotal column data (like TOMS and GOME) or data with higher vertical resolution (like SBUV).

5.7.5 Scatterometer winds

In HOP, the observation array is scanned for SCAT data. Normally two ambiguous pairs of u-component and v-component observations are found at each SCAT location – with directions approximately 180 degrees apart. QuikSCAT can have 2, 3 or 4 ambiguous winds. Only the most likely wind (highest *a priori* probability) and its opposing (135 to 180 apart) ambiguity are used. If there are no ambiguities in the allowed range then all ambiguities are rejected. If LQSCATT = .TRUE. (the default, modifiable through the namelist NAMJO), the normal quadratic J_o will be used. In this case only the SCAT wind nearest to the high resolution background will be used (which is determined in a section of HOP). Otherwise, both winds are used and the ambiguity removal takes place implicitly through a special SCAT cost-function, (2.14), in HJO (Stoffelen and Anderson, 1997).

As PPUV10M (Section 5.8) is used also for SCAT data (since Cy18r6), the observation operator is exactly the same as for SYNOP. SHIP and DRIBU winds. The z_0 (surface roughness) comes from the coupled wave model. The simpler logarithmic wind law can be used optionally under the switch LSCASUR = .FALSE. in NAMOBS (.TRUE. by default).

In the adjoint (SURFACAD) there is a separate section of HOP for the calculation of the $\nabla_{\mathrm{obs}} J_{\mathrm{SCAT}}$.

5.7.6 GPS Radio Occultation bending angles

The subroutine GPSRO_OP is called in HOP and it simulates GPS radio occulation bending angles using the one-dimensional model outlined in Healy and Thépaut (2006). The subroutine evaluates a profile of bending angles, α as function of impact parameter, a, at each observation location by evaluating the integral

$$\alpha(a) = -2a \int_a^\infty \frac{\frac{d\ln n}{dx}}{(x2 - a2)^{1/2}} dx \tag{5.34}$$

where n is the refractive index and $x = nr$, the product of the refractive index and r, a radial coordinate value. The pressure, temperature, specific humidity and geopotential on the model levels (ZPRESF5, ZTF5, ZQF5 and ZGEOPF5, respectively) produced by PREINT (see Subsection 5.5.1) are the inputs to GPSRO_OP. The observation operator calculates the refractivity (defined as $N = 10^{-6}(n-1)$) on the full model levels using the pressure, temperature and specific humidity profiles. It then converts the geopotential heights to geometric heights and then radius values. The bending angle integral is evaluated assuming that the refractivity, N, varies exponentially between the model levels.

The bending angle observation errors are set in GPSRO_OBERROR, which is called in HRETR. Entries in the JO tables are set in SULIMB.

The routine GPSRO_OP is used when local horizontal homogeneity is to be assumed for the bending angle computations. Alternatively, the ray-tracing can take the horizontal structure in the limb-viewing plane into account by providing a series of profiles covering the limb-viewing plane. In this case, profiles provided by PREINT2D are used, and the routine GPSRO_2DOP is called from HOP instead of GPSRO_OP. The 2d facility is switched on by specifying NOBSPROFS(10) > 1 for obstype 10 in the namelist NAMNPROF. Note that this means any limb radiances present in the assimilation will also take horizontal structure into account.

5.8 SURFACE OBSERVATION OPERATORS

All surface data are processed in the routine SURFACEO Preparations for the vertical interpolation is done as for all other data in PREINT (see Subsection 5.5.1), and for surface data there are a few additional tasks which are performed in a separate routine, PREINTS. In PREINTS. surface roughness over sea, dry static energy (SURBOUND), Richardson number, drag coefficients and stability functions (EXCHCO), are computed, as detailed in the following.

5.8.1 Mathematical formulation

An analytical technique (Geleyn, 1988) is used to interpolate values between the lowest model level and the surface. Using Monin–Obukhov theory gives

$$\frac{\partial u}{\partial z} = \frac{u_*}{\kappa(z+z_0)}\phi_{\mathrm{M}}\left(\frac{z+z_0}{L}\right) \tag{5.35}$$

$$\frac{\partial s}{\partial z} = \frac{s_*}{\kappa(z+z_0)}\phi_{\mathrm{H}}\left(\frac{z+z_0}{L}\right) \tag{5.36}$$

$$L = \frac{c_p}{g}\frac{T}{\kappa}\frac{u_*^2}{s_*} \tag{5.37}$$

where u, s are wind and energy variables,u_*, s_* are friction values and $\kappa = 0.4$ is von Kármán's constant.

The temperature is linked to the dry static energy s by

$$s = c_p T + \phi \tag{5.38}$$

$$c_p = c_{p_{\mathrm{dry}}}\left[1 + \left(\frac{c_{p_{\mathrm{vap}}}}{c_{p_{\mathrm{dry}}}} - 1\right)q\right] \tag{5.39}$$

Defining the neutral surface exchange coefficient at the height z as

$$C_{\mathrm{N}} = \left[\frac{\kappa}{\ln\left(\frac{z+z_0}{z_0}\right)}\right]^2 \tag{5.40}$$

The drag and heat coefficients as

$$C_{\mathrm{M}} = \frac{u_*^2}{[u(z)]^2} \tag{5.41}$$

$$C_{\mathrm{H}} = \frac{u_* s_*}{u(z)[s(z) - \tilde{s}]} \tag{5.42}$$

we can set the following quantities

$$B_{\mathrm{N}} = \frac{\kappa}{\sqrt{C_{\mathrm{N}}}}, \quad B_{\mathrm{M}} = \frac{\kappa}{\sqrt{C_{\mathrm{M}}}}, \quad B_{\mathrm{H}} = \frac{\kappa\sqrt{C_{\mathrm{M}}}}{C_{\mathrm{H}}} \tag{5.43}$$

and considering the stability function in stable conditions as

$$\phi_{\mathrm{M/H}} = 1 + \beta_{\mathrm{M/H}}\frac{z}{L} \tag{5.44}$$

we obtain by integrating (5.35) and (5.36) from 0 to z_1 (the lowest model level) values of $u(z)$ and $s(z)$ given by

$$u(z) = \frac{u(z_1)}{B_{\mathrm{M}}}\left[\ln\left(1 + \frac{z}{z_1}(e^{B_{\mathrm{N}}} - 1)\right) - \frac{z}{z_1}(B_{\mathrm{N}} - B_{\mathrm{M}})\right] \tag{5.45}$$

$$s(z) = \tilde{s} + \frac{s(z_1) - \tilde{s}}{B_{\mathrm{H}}}\left[\ln\left(1 + \frac{z}{z_1}(e^{B_{\mathrm{N}}} - 1)\right) - \frac{z}{z_1}(B_{\mathrm{N}} - B_{\mathrm{H}})\right] \tag{5.46}$$

In unstable conditions the stability function can be expressed as

$$\phi_{M/H} = \left(1 - \beta_{M/H}\frac{z}{L}\right)^{-1} \tag{5.47}$$

and the vertical profiles for wind and dry static energy are

$$u(z) = \frac{u(z_1)}{B_M}\left[\ln\left(1 + \frac{z}{z_1}(e^{B_N} - 1)\right) - \ln\left(1 + \frac{z}{z_1}(e^{B_N - B_M} - 1)\right)\right] \tag{5.48}$$

$$s(z) = \tilde{s} + \frac{s(z_1) - \tilde{s}}{B_H}\left[\ln\left(1 + \frac{z}{z_1}(e^{B_N} - 1)\right) - \ln\left(1 + \frac{z}{z_1}(e^{B_N - B_H} - 1)\right)\right] \tag{5.49}$$

The temperature can then be obtained from s as

$$T(z) = s(z) - \frac{zg}{c_p} \tag{5.50}$$

When z is set to the observation height, (5.45), (5.46) and (5.48) to (5.50) give the postprocessed wind and temperature. To solve the problem, we have to compute the dry static energy at the surface $\tilde{s} = \tilde{s}(T_{surf}, q = 0)$ (Subsection 5.8.2), with B_M, B_N, and B_H values depending on the drag and heat exchange coefficients as detailed in Subsection 5.8.3.

5.8.2 Surface values of dry static energy

To determine the dry static energy at the surface we use (5.38) and (5.39) where the humidity at the surface is defined by

$$\tilde{q} = q(z = 0) = h(C_{snow}, C_{liq}, C_{veg})q_{sat}(T_{surf}, p_{surf}) \tag{5.51}$$

where, according to Blondin (1991), h is given by

$$h = C_{snow} + (1 - C_{snow})[C_{liq} + (1 - C_{liq})\bar{h}] \tag{5.52}$$

with

$$\bar{h} = \max\left\{0.5\left(1 - \cos\frac{\pi\vartheta_{soil}}{\vartheta_{cap}}\right), \min\left(1, \frac{q}{q_{sat}(T_{surf}, p_{surf})}\right)\right\} \tag{5.53}$$

where ϑ_{soil} is the soil moisture content and ϑ_{cap} is the soil moisture at field capacity (2/7 in volumetric units). Equation (5.52) assigns a value of 1 to the surface relative humidity over the snow covered and wet fraction of the grid box. The snow-cover fraction C_{snow} depends on the snow amount W_{snow} so that

$$C_{snow} = \min\left(1, \frac{W_{snow}}{W_{snow_{cr}}}\right)$$

where $W_{snow_{cr}} = 0.015$ m is a critical value. The wet skin fraction C_{liq} is derived from the skin-reservoir water content W_{liq} by

$$C_{liq} = \min\left(1, \frac{W_{liq}}{W_{liq_{max}}}\right),$$

where

$$W_{liq_{max}} = W_{layer_{max}}\{(1 - C_{veg}) + C_{veg}A_{leaf}\}$$

with $W_{layer_{max}} = 2 \times 10^{-4}$ m being the maximum amount of water that can be held on one layer of leaves, or as a film on bare soil, $A_{leaf} = 4$ is the leaf-area index, and C_{veg} is the vegetation fraction.

5.8.3 Transfer coefficients

Comparing the (5.35) and (5.36) integrated from z_o to $z + z_0$ with (5.40) to (5.42) , C_M and C_H can be analytically defined:

$$\frac{1}{C_M} = \frac{1}{\kappa^2}\left[\int_{z_0}^{(z+z_0)} \frac{\phi_M(z'/L)}{z'}\,dz'\right]^2 \tag{5.54}$$

$$\frac{1}{C_H} = \frac{1}{\kappa^2}\left[\int_{z_0}^{(z+z_0)} \frac{\phi_M(z'/L)}{z'}\,dz' \int_{z_0}^{(z+z_0)} \frac{\phi_H(z'/L)}{z'}\,dz'\right] \tag{5.55}$$

Because of the complicated form of the stability functions, the former integrals have been approximated by analytical expressions, formally given by

$$\begin{aligned} C_{\mathrm{M}} &= C_{\mathrm{N}} f_{\mathrm{M}}\left(Ri, \frac{z}{z_0}\right) \\ C_{\mathrm{H}} &= C_{\mathrm{N}} f_{\mathrm{H}}\left(Ri, \frac{z}{z_0}\right) \end{aligned} \tag{5.56}$$

where C_{N} is given by (5.40). The bulk Richardson number Ri is defined as

$$Ri = \frac{g\Delta z \Delta T_{\mathrm{v}}}{c_p T_{\mathrm{v}} |\Delta \underline{u}|^2} \tag{5.57}$$

where T_{v} is the virtual potential temperature. The functions f_{M} and f_{H} correspond to the model instability functions and have the correct behaviour near neutrality and in the cases of high stability (Louis, 1979; Louis *et al.*, 1982).

(i) *Unstable case* $Ri < 0$

$$f_{\mathrm{M}} = 1 - \frac{2bRi}{1 + 3bCC_{\mathrm{N}}\sqrt{\left(1 + \frac{z}{z_0}\right)(-Ri)}} \tag{5.58}$$

$$f_{\mathrm{H}} = 1 - \frac{3bRi}{1 + 3bCC_{\mathrm{N}}\sqrt{\left(1 + \frac{z}{z_0}\right)(-Ri)}} \tag{5.59}$$

with $C = 5$

(ii) *Stable case* $Ri > 0$

$$f_{\mathrm{M}} = \frac{1}{1 + 2bRi(1 + dRi)^{1/2^{1/2}}} \tag{5.60}$$

$$f_{\mathrm{H}} = \frac{1}{1 + 3bRi(1 + dRi)^{1/2^{1/2}}} \tag{5.61}$$

with $d = 5$

5.8.4 Two-metre relative humidity

In GPRH relative humidity is computed according to (5.28). The relative humidity depends on specific humidity, temperature and pressure (q, T and p, respectively) at the lowest model level. It is constant in the surface model layer, see PPRH2M.

Chapter 6
Background, analysis and forecast errors

Table of contents

6.1 NOMENCLATURE

The calculation of standard deviations of background errors is unfortunately an area of confusing nomenclature. "Standard deviations of background error" is quite a mouthful, so they are generally referred to simply as 'background errors' (likewise for standard deviations of analysis errors and forecast errors). Although inaccurate, this nomenclature has been adopted in the following for the sake of brevity.

A second source of confusion is that the terms 'background error' and 'forecast error' are often used interchangeably. This confusion has even crept into the code, where the buffer which contains the standard deviations of background error is called FCEBUF. Such confusion is clearly unwise when discussing their calculation. In the following, we describe the processing of error variances during a single analysis cycle. The term 'background error' will be used exclusively to refer to the standard deviations of background error used in the background cost function. The background errors are an input to the analysis. The term 'forecast error' will refer to an estimate of the standard deviation of error in a short-term forecast made from the current analysis. The forecast errors are calculated by inflating an estimate of the standard deviation of analysis error, and are an output from the analysis system. They provide background errors for the *next* analysis cycle.

6.2 INPUT AND 'MASSAGING' OF BACKGROUND ERRORS

Background errors for use in J_b are initialised by a call to SUINFCE. This is part of the J_b set-up, which is described in Subsection 4.3.3.

As with most of the J_b code, the structure SPJB_VARS_INFO, described in Subsection 4.3.1, is important. At the beginning of SUINFCE, this structure is examined to determine which variables are present in J_b, and to locate the corresponding profiles of globally-averaged background error. Next, a call to IO_INQUIRE is made to determine which fields are present in the input file (filename **errgrib**), and to find details of their grid resolution, etc.

The fields of background error are read by a call to IO_GET, and interpolated to the model's horizontal grid using SUHIFCE.

Vertical interpolation requires knowledge of the backgound surface pressure. In addition, one method of constructing humidity background errors requires knowledge of background values of specific humidity and temperature. So, background fields in grid space are obtained either by a call to GET_TRAJ_GRID, or by transforming the corresponsing spectral fields.

Before the vertical interpolation takes place, the locations of the different J_b variables within the background error array FCEBUF are found by a set of calls to the internal subroutine LOOKUP, which interrogates SPJB_VARS_INFO.

The large loop "VARIABLE_LOOP" in SUINFCE loops over the fields of the input file. Each field is interpolated onto model levels by a call to SUVIFCE (unless it is a surface field), and either stored in FCEBUF or used to construct a field in FCEBUF. (For example, if backgound errors for vorticity are not available in the input file, they are constructed by scaling zonal wind errors.)

Background errors for the humidity variable are treated as a special case, and generated after the main loop over variables by a call to SUSHFCE. The calculation of background errors for humidity is described in Subsection 4.3.4.

Next, one of two routines is called. SUMDFCE calculates a vertically averaged 'pattern' of background error. This is required if the background errors are to be represented as a product of a vertical profile of global mean error and a horizontal pattern. The pattern is stored in FGMWNE. (Note in particular that SUMDFCE is called if horizontally-constant background errors are requested by setting LCFCE. In this case, all elements of FGMWNE are set to one.)

Alternatively, SUPRFFCE is called to calculate global mean profiles of the input background errors. This is the default. The profiles are stored in FCEIMN.

The final step in processing the background errors is performed by SUSEPFCE. This modifies the background errors in one of two ways. If separable background errors have been requested, the contents of the background error buffer are replaced by the product of the vertical profile stored in FCEMN and the horizontal pattern stored in FGMWNE. Otherwise, the background errors for each variable at each level are multiplied by the ratio of the corresponding elements of FCEMN and FCEIMN. The result of this operation is to adjust the global mean profiles of background error to match those stored in FCEMN.

6.3 DIAGNOSIS OF BACKGROUND-ERROR VARIANCES

Analysis errors are calculated by subtracting a correction from the variances of background error. The first stage in the calculation is therefore to determine the background-error variances. This is done by subroutine BGVECS, which is called from FORECAST_ERROR. One of two methods may be employed, depending on whether NBGVECS is equal to, or greater than, zero. In either case, the estimated variances of background error are stored in the analysis error buffer, ANEBUF (in YOMANEB).

If NBGVECS is zero, then background errors for variables which are explicitly present in the background error buffer, FCEBUF, are copied into ANEBUF and squared. Errors for those variables whose background errors are defined implicitly through the change of variable are estimated using simple scaling of appropriate explicit errors. This scaling is performed by a call to ESTSIGA.

If NBGVECS is non-zero, then the variances of background error are estimated using randomization. This method assumes that the change of variable transforms the background error covariance matrix into the identity matrix. A sample of NBGVECS vectors drawn from a multi-dimensional Gaussian distribution with zero mean and identity covariance matrix is generated by calls to RANDOM_CTLVEC. These vectors are transformed to the space of physical variables by CHAVARIN. The transformed variables form a sample drawn from the distribution of background errors. A call to STEPOTL('0AA00A000') transforms each vector to gridpoint space and accumulates the sums of squares in ANEBUF. In addition, observation operators are appled to the random variables to generate samples of background error in terms of radiance, for example. The sums of squares of these observation-space variables are also accumulated to provide estimates of background error in observation space.

Finally, the sums of squares of random vectors are divided by the number of vectors by a call to SCALEAE to provide a somewhat noisy estimate of the variances of background error actually used in the analysis. Noise may be filtered by a call to FLTBGERR, which transforms the variances to spectral coefficients, multiplies each coefficient by $\cos^2(\min((n/\text{NBGTRUNC}), 1)\pi/2)$, and then transforms to grid space. The default is to filter with a very large value of NBGTRUNC. Effectively, the background errors are simply

spectrally truncated. It is highly recommended that the filtering is performed, since it prevents a grid-scale numerical instability which occurs when the error growth model introduces spatial features which cannot be resolved by the spectral control variable.

Flow-dependent background errors, valid at the end of the window may be generated by setting LBGM (namvar) to TRUE. In this case, the tangent linear model is used to propagate each of the random background vectors to the end of the analysis window. The eigenvectors of the analysis Hessian (see the next section) are also propagated in time, by a call to CNT3TL from XFORMEV. For the normal case, LBGM is false, and background and analysis errors are estimated at the beginning of the analysis window, and propagated using the simple error-growth model described in section 6.7.

The background errors diagnosed by BGVECS may be written out for diagnostic purposes by setting LWRISIGB. The errors are written by a call to WRITESD (called from FORECAST_ERROR).

6.4 CALCULATION OF EIGENVALUES AND EIGENVECTORS OF THE HESSIAN

The second stage in the calculation of analysis errors is to determine eigenvalues and eigenvectors of the Hessian of the cost function. This is done using a combined Lanczos and conjugate-gradient algorithm, CONGRAD, called from CVA1 under the control of LAVCGL. The reader is referred to Fisher (1998, ECMWF Seminar proceedings pp364-385) for a detailed description of the CONGRAD algorithm. Note that CONGRAD requires that the cost function is strictly quadratic.

CONGRAD starts by transforming the initial control variable and gradient to a space with Euclidian inner product. Typically, this transformation is simply a multiplication by YRSCALPSQRT, but may also involve preconditioning via calls to PRECOND. The transformed initial gradient is normalized to give the first Lanczos vector. The Lanczos vectors are stored in the array YL_ZCGLWK.

Each iteration of the conjugate-gradient/Lanczos algorithm starts by calculating the product of the Hessian and the latest search direction. This is calculated as $J''\mathrm{d} = [\nabla J(\mathrm{x}_0 + \mathrm{d}) - \nabla J(\mathrm{x}_0)]$, where d" is a vector of unit length. This finite difference formula is exact, since the cost function is quadratic.

The main iteration loop calculates the sequence of gradients, and the sequence of coefficients according to the the Lanczos recurrence. The sequence of control vectors that partially minimize the cost function is not explicitly generated. Thus, unlike e.g. M1QN3, the control vectors passed to SIM4D do not lie on a path towards the minimum, and the gradients returned by SIM4D do not, in general, decrease as the minimization proceeds. The optimal control vector and the corresponding gradient can, however, be determined as a linear combination of the Lanczos vectors.

In general, only the gradient is calculated at each iteration. It is required in order to monitor the convergence of the minimization. Optionally, by setting L_CHECK_GRADIENT, the optimal point is also calculated, and an additional call to SIM4D is made to evaluate the true gradient at the optimal point. This is useful as a means of checking the assumption of linearity on which the algorithm is based. However, it doubles the cost of the minimization.

The Lanczos algorithm produces a sequence of coefficients, which are stored in the arrays ZDELTA and ZBETA. These correspond to the diagonal and sub-diagonal of a symmetric tri-diagonal matrix. The calculation of the optimal point and the corresponding gradient require the solution of a linear system involving this matrix. This is performed by the internal subroutine PTSV, which is a simple interface to the LAPACK routine SPTSV. In addition, the eigenvalues of the tri-diagonal matrix are approximations to eigenvalues of the Hessian of the cost function. These approximate eigenvalues are calculated every iteration, together with bounds on their accuracy. As soon as the leading eigenvalue has converged sufficiently, it is monitored to check that it does not increase. This provides a sensitive test that the algorithm is behaving correctly. Any increase in the leading eigenvalue provides an early indication of failure (for example, due to a bad gradient) and the algorithm is immediately terminated. The calculation is not aborted, since the test detects the failure of the algorithm before the converged eigenvalues and eigenvectors become corrupted.

A call to PREDICT_RUNTIME is made every iteration. This subroutine attempts to predict how many iterations will be required to minimize the cost function, and how long this is likely to take. These predictions are written to STDERR.

After the last iteration, the converged eigenvectors of the Hessian are calculated by calling WREVECS. Note that the criterion used to decide which eigenvalues have converged is relaxed at this stage to $\|J''\mathbf{v} - \lambda\mathbf{v}\| < \varepsilon\|\mathbf{v}\|$, where ε is given by EVBCGL. The default value for EVBCGL is 0.1.

Next, if required, upper and lower bounds for $N - p^{\mathrm{T}}(J")^{-1}p$ (where N is the dimension of the control vector) and $-p^{\mathrm{T}}\log_2[(J")^{-1}]p$ are calculated using the algorithm of Golub and Meurant (1994). These quantities may be used to evaluate the information content of the analysis using the switch L_INFO_CONTENT.

Finally, CONGRAD calculates the the optimal control vector and gradient as a linear combination of the Lanczos vectors, and transforms them from the Euclidian space used internally to the usual space of the control variable.

6.5 THE PRECONDITIONER

CONGRAD allows the use of a preconditioner. The preconditioner is a matrix which approximates the Hessian matrix of the cost function. The preconditioner used in CONGRAD is a matrix of the form

$$\mathbf{I} + \sum_{i=1}^{L}(\mu_i - 1)\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \tag{6.1}$$

where the vectors $\mathbf{w}_i$ are orthogonal. The pairs $\{\mu_i, \mathbf{w}_i\}$ are calculated in PREPPCM, and are intended to approximate some of the eigenpairs (i.e. eigenvalues and associated eigenvectors) of the Hessian matrix of the cost function. They are calculated as follows.

A set of L vectors, $\mathbf{u}_i$, is read in using READVEC. These vectors may be in the space of the control vector (if LEVECCNTL is true), or in model space (if LEVECCNTL is false). In the latter case, the vectors are transformed to control space by calls to CHAVAR.

The vectors (transformed, if necessary to control space) are assumed to satisfy

$$\mathbf{I} - \sum_{i=1}^{L}\mathbf{u}_i\mathbf{u}_i^{\mathrm{T}} \approx (J'')^{-1} \tag{6.2}$$

Vectors which meet this criterion can be written out from an earlier forecast error calculation by setting LWRIEVEC.

The input vectors are not necessarily orthogonal, whereas the preconditioner requires a set of orthogonal vectors. Let us denote by $\mathbf{U}$ the matrix whose columns are the vectors $\mathbf{u}_i$. A sequence of Householder transformations is now performed to transform $\mathbf{U}$ to upper triangular. Let us represent this sequence of Householder transformations by the matrix $\mathbf{Q}$. Then $\mathbf{QU}$ is upper triangular, which means that $(\mathbf{QU})(\mathbf{QU})^{\mathrm{T}}$ is zero except for an $L \times L$ block in the top left hand corner.

It is clear that $(\mathbf{QU})(\mathbf{QU})^{\mathrm{T}}$ has only L non-zero eigenvalues. Moreover, the non-zero eigenvalues are the eigenvalues of the $L \times L$ block matrix, and the eigenvectors of $(\mathbf{QU})(\mathbf{QU})^{\mathrm{T}}$ are the eigenvectors of the block matrix, appended by zeroes. These eigenvalues and eigenvectors are calculated by a call to the LAPACK routine SSYEV.

Now, since $\mathbf{Q}$ is an orthogonal matrix, we have $\mathbf{QQ}^{\mathrm{T}} = \mathrm{I}$. So, we may write (6.2) as

$$\mathbf{I} - \mathbf{Q}^{\mathrm{T}}(\mathbf{QU})(\mathbf{QU})^{\mathrm{T}}\mathbf{Q} \approx (J'')^{-1} \tag{6.3}$$

Let us denote the eigenpairs of $(\mathbf{QU})(\mathbf{QU})^{\mathrm{T}}$ by $\{\rho_i, \mathbf{v}_i\}$. Then we may write (6.3) as

$$\mathbf{I} - \sum_{i=1}^{L}\rho_i(\mathbf{Q}^{\mathrm{T}}\mathbf{v}_i)(\mathbf{Q}^{\mathrm{T}}\mathbf{v}_i)^{\mathrm{T}} \approx (J'')^{-1} \tag{6.4}$$

The orthogonality of $\mathbf{Q}$ and the orthonormality of the eigenvectors $\mathbf{v}_i$, means that the vectors $\mathbf{Q}^{\mathrm{T}}\mathbf{v}_i$ are orthonormal. They are, in fact, the required vectors, $\mathbf{w}_i$ of the preconditioner matrix.

Inverting (6.4) gives

$$\mathbf{I} - \sum_{i=1}^{L} \frac{1}{\rho_i}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx J'' \tag{6.5}$$

Defining $\mu_i = 1 - 1/\rho_i$ gives the required approximation to the Hessian matrix.

The preconditioner vectors are sorted in decreasing order of μ_i, and all vectors for which $\mu_i < 1$ are rejected. These vectors cannot be good approximations to eigenvectors of the Hessian matrix, since the eigenvalues of the Hessian matrix are all greater than or equal to one. A final refinement to the calculation is to reduce large values of μ_i to a maximum of R_MAX_CNUM_PC (typically 10). This was found to be necessary in practice to avoid ill-conditioning the minimization.

The numbers μ_i are stored in RCGLPC. The vectors, $\mathbf{w}_i$ are stored in YVCGLPC, and the total number of preconditioner vectors is stored in NVCGLPC.

Application of the preconditioner is straightforward, and is performed by subroutine PRECOND. This routine can also apply the inverse, the symmetric square root, or the inverse of the symmetric square root of the preconditioner matrix. Application of the latter matrices relies on the observation that if

$$\mathbf{M} = \mathbf{I} + \sum_{i=1}^{L}(\mu_i - 1)\mathbf{w}_i\mathbf{w}_{\mathrm{i}}^{\mathrm{T}} \tag{6.6}$$

with orthonormal $\mathbf{w}_i$, then the expressions for $\mathbf{M}^{-1}$, $\mathbf{M}^{1/2}$ and $\mathbf{M}^{-1/2}$ result from replacing μ_i in (6.6) by $1/\mu_i$, $\sqrt{\mu_i}$ and $1/(\sqrt{\mu_i})$ respectively.

6.6 CALCULATION OF ANALYSIS-ERROR VARIANCES

The eigenvectors and eigenvalues of the Hessian matrix calculated by CONGRAD are passed to XFORMEV, which uses them to estimate the analysis-error variances.

The first step is to undo any preconditioning. If preconditioning has been employed, then the eigenvectors and eigenvalues produced by CONGRAD provide an approximation to the preconditioned Hessian, $\mathbf{M}^{-1/2}J''\mathbf{M}^{-1/2}$, of the form

$$\mathbf{M}^{-1/2}J''\mathbf{M}^{-1/2} \approx \mathbf{I} + \sum_{i=1}^{K}(\lambda_i - 1)\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}} \tag{6.7}$$

Multiplying to the left and right by $\mathbf{M}^{1/2}$, gives

$$J'' \approx \mathbf{M} + \sum_{i=1}^{K}(\lambda_i - 1)(\mathbf{M}^{1/2}\mathbf{v}_i)(\mathbf{M}^{1/2}\mathbf{v}_i)^{\mathrm{T}} \tag{6.8}$$

Substituting for the preconditioner matrix from (6.6), gives the following

$$J'' \approx \mathbf{I} + \sum_{i=1}^{L+K} \mathbf{s}_i\mathbf{s}_i^{\mathrm{T}} \tag{6.9}$$

where

$$\mathbf{s}_i = \begin{cases} (\mu_i - 1)^{1/2}\mathbf{w}_i & \text{for } i = 1 \ldots L \\ (\lambda_{i-L} - 1)^{1/2}\mathbf{M}^{1/2}\mathbf{v}_{i-L} & \text{for } i = L+1 \ldots L+K \end{cases} \tag{6.10}$$

Note that the resulting approximation to the Hessian is not expressed in terms of eigenvalues and eigenvectors. Consequently, inversion of the approximation must be performed using the Shermann–Morrison–Woodbury formula. Let $\mathbf{S}$ be the matrix whose columns are the vectors $\mathbf{s}_i$. Then, according to

the Shermann–Morrison–Woodbury formula, the inverse of the approximate Hessian matrix is

$$(J'')^{-1} \approx \mathbf{I} - \mathbf{S}(\mathbf{I} + \mathbf{S}^{\mathrm{T}}\mathbf{S})^{-1}\mathbf{S}^{\mathrm{T}} \tag{6.11}$$

The matrix $(\mathbf{I} + \mathbf{S}^{\mathrm{T}}\mathbf{S})$ has dimension $(L + K) \times (L + K)$. This matrix is constructed, and its Cholesky decomposition is calculated using the LAPACK routine SPOTRF. This gives a lower triangular matrix $\mathbf{C}$ such that

$$(J'')^{-1} \approx \mathbf{I} - (\mathbf{S}\mathbf{C}^{-1})(\mathbf{S}\mathbf{C}^{-1})^{\mathrm{T}} \tag{6.12}$$

The matrix $(\mathbf{S}\mathbf{C}^{-1})$ is calculated by back-substitution.

The final stage in the calculation of the analysis errors is to transform the columns of the matrix $(\mathbf{S}\mathbf{C}^{-1})$ to the space of model variables by applying the inverse change of variable, CHAVARIN. This gives the required approximation to the analysis error covariance matrix

$$\mathbf{P}_{\mathrm{a}} \approx \mathbf{B} - \mathbf{V}\mathbf{V}^{\mathrm{T}} \tag{6.13}$$

where $\mathbf{V} = \mathbf{L}^{-1}\mathbf{S}\mathbf{C}^{-1}$, and where $\mathbf{L}^{-1}$ represents the inverse of the change of variable. The columns of $\mathbf{V}$ may be written out (e.g. for diagnostic purposes, or to form the preconditioner for a subsequent minimization) by setting LWRIEVEC. The columns of $\mathbf{V}$ are then transformed to gridpoint space, and their sums of squares (i.e. the diagonal elements of $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ in gridpoint space) are subtracted from the variances of background error which were stored in ANEBUF before the minimization by BGVECS.

The analysis errors are calculated as the difference between the background errors and a correction derived from the eigenvectors of the Hessian. If the background errors are underestimated, there is a danger that the correction will be larger than the background error, giving negative variances of analysis error. This is unlikely to happen if the background errors are estimated using randomization, or for variables whose background errors are explicitly specified in the background cost function, but is possible for variables such as temperature whose background errors are not explicitly specified. To guard against this eventuality, if NBGVECS is zero, then the variances of analysis error for variables whose background errors are not explicit are estimated by applying a scaling to the explicit variables by a call to ESTSIGA from CVA1. The variances are then converted to standard deviations and written out by a call to WRITESD.

6.7 CALCULATION OF FORECAST-ERROR VARIANCES

The analysis errors are inflated according to the error growth model of Savijärvi (1995) to provide estimates of short-term forecast error. This is done by a call to ESTSIG.

The error growth model is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = (a + b\sigma)\left(1 - \frac{\sigma}{\sigma_\infty}\right) \tag{6.14}$$

Here, a represents growth due to model errors, b represents the exponential growth rate of small errors, and σ_∞ represents the standard deviation of saturated forecast errors. The growth due to model error is set to 0.1 times the global mean background error per day. The exponential growth rate, b, is set to 0.4 per day.

The saturation standard deviations are calculated as $\sqrt{2}$ times the standard deviation of each field. The standard deviations have been calculated for each month from the re-analysis dataset. ESTSIG reads these climatological error fields from file 'stdev_of_climate' by calling READGRIB, and interpolates them in the horizontal and vertical using SUHIFCE and SUVIFCE. The climatological errors may also be artificially increased in the tropics under the control of LFACHRO. If climate standard deviations are not available for any field, they are estimated as 10 times the global mean background error for the field.

The error growth model is integrated for a period of NFGFCLEN hours. The integration is done analytically using the expression given by Savijärvi (1995). Two precautions are taken in integrating the error growth model. First, negative analysis-error variances are set to zero. Second, the growth rate due to model error is limited to a sensible value with respect to the saturation errors. This was found to be necessary to prevent numerical problems when calculating specific humidity errors for the upper levels of the model.

ESTSIG overwrites the contents of ANEBUF with the estimated variances of forecast error. The variances are converted to standard deviations and written out by WRITESD.

Chapter 7
Gravity-wave control

Table of contents

7.1 INTRODUCTION

In 3D-Var, gravity-wave control is achieved via a penalty term J_c based on the techniques of normal-mode initialization (NMI), in 4D-Var a weak constraint digital filter is used.

Section 7.2 provides a brief overview of NMI techniques, together with references to scientific papers in which further details can be found. Section 7.3 describes the computation of normal modes and related arrays. Section 7.4 documents the implementation of nonlinear NMI in 3D- and 4D-Var, while Section 7.5 describes the computation of J_c based on NMI. Section 7.6 gives an overview of digital filter initialization techniques while Section 7.7 describes its implementation as it is used in the 4D-Var assimilation system.

7.2 NORMAL-MODE INITIALIZATION

If the model equations are linearized about a state of rest, the solutions can (with a certain amount of arbitrariness) be classified into 'slow' (Rossby) and 'fast' (gravity) modes. This classification defines two mutually orthogonal subspaces of the finite-dimensional vector space containing the model state $\mathbf{x}$. Thus, the model state can be written as

$$\mathbf{x} = \mathbf{x}_{\mathrm{R}} + \mathbf{x}_{\mathrm{G}} \tag{7.1}$$

where $\mathbf{x}_{\mathrm{R}}$ is the 'slow' component and $\mathbf{x}_{\mathrm{G}}$ the 'fast' component. *Linear* NMI consists of removing the fast component altogether ($\mathbf{x}_{\mathrm{G}} = 0$). Since the model is nonlinear, a much better balance is obtained by setting the *tendency* of the fast component to zero ($\dot{\mathbf{x}}_{\mathrm{G}} = 0$); it is this balance condition which *nonlinear* NMI seeks to impose.

Nonlinear NMI was first demonstrated by Machenhauer (1977), in the context of a spectral shallow-water model. For a multi-level model, the first stage in the modal decomposition is a vertical transform; each vertical mode then has its own set of horizontal slow and fast modes (for the shallower vertical modes, all the corresponding horizontal modes can be considered as 'slow'). In the case of a multi-level spectral model using the ECMWF hybrid vertical coordinate the details may be found in the report by Wergen (1987), which also describes techniques for taking into account forcing by physical (non-adiabatic) processes and the diurnal and semi-diurnal tidal signals. Although these options are still coded in the IFS, they are no longer used operationally at ECMWF and will not be described in this documentation.

Implicit normal mode initialization (Temperton, 1988) is based on the observation that, except at the largest horizontal scales, the results of NMI can be reproduced almost exactly without computing the

horizontal normal modes at all. The calculation reduces to solving sets of elliptic equations. In the case of a spectral model (Temperton, 1989), these sets of equations are tridiagonal in spectral space. The IFS code includes the option of 'partially implicit NMI', in which the initialization increments are computed using the full 'explicit' NMI procedure for large horizontal scales while the remaining increments at smaller horizontal scales are computed using the simpler implicit procedure.

7.3 COMPUTATION OF NORMAL MODES

7.3.1 Vertical modes

The vertical normal modes depend on the number of levels in the model and on their vertical distribution. They also depend on the choice of reference temperature SITR (assumed isothermal) and reference surface pressure (SIPR). The vertical modes used by the initialization routines are also used in the semi-implicit scheme for the forward integration of the model. The computation of J_b and J_c also uses the vertical normal modes, but for these purposes different values of SITR and SIPR may be selected. Thus the vertical modes are computed both in SUDYN and SUSINMI, the latter being used especially in 4D-Var where it is necessary to alternate between applications using different choices of SITR and SIPR. The vertical modes are computed by first calling SUBMAT to set up a vertical structure matrix and then calling an eigenvalue/eigenvector routine EIGSOL (at the end of SUDYN, it calls routine RG in the auxiliary library). After reordering and normalization, the eigenvectors (vertical modes) are stored in the matrix SIMO, while the corresponding eigenvalues (equivalent depths) are stored in the array SIVP. The inverse of SIMO is computed and stored in SIMI.

7.3.2 Horizontal modes and help arrays

The horizontal normal modes depend on the equivalent depths (see above) and the chosen spectral truncation NXMAX. For 'explicit' NMI, NXMAX is equal to the model's spectral truncation NSMAX. Normally, 'partially implicit NMI' is chosen by setting the switch LRPIMP to .TRUE. In this case the explicit NMI increments are used only up to spectral truncation NLEX (21 by default) but in order to blend the explicit and implicit increments smoothly, explicit increments are computed up to a slightly higher resolution. By default, NXMAX = NLEX + 5.

For most applications of the NMI procedure in the operational suite, it is considered that the larger horizontal scales are best left uninitialized (they include, for example, atmospheric tidal signals and large-scale tropical circulations driven by diabatic processes). To cater for this option there is another logical switch, LASSI ('adiabatic small-scale initialization'), which sets to zero all the initialization increments for total wavenumbers up to NFILTM (= 19 by default). Since only the small-scale increments are used, the NMI can be completely implicit: NLEX is set to 0 and there is no need to calculate the 'explicit' horizontal normal modes.

All the horizontal-normal-mode computations are carried out only for the first NVMOD vertical modes. By default, NVMOD = 5.

The horizontal modes are computed by calling SUMODE3. In turn, SUMODE3E computes the explicit modes and their frequencies while SUMODE3I computes the 'help' arrays required to invert the tridiagonal systems encountered in implicit NMI.

7.4 IMPLEMENTATION OF NMI

Nonlinear NMI can be invoked by calling NNMI3. This is no longer applied by default in IFS, now all gravity-wave control during the assimilation process is done through the penalty term or digital filter. But signicant parts of the NMI code is still used for this, as described in Section 7.5. Model tendencies are computed by calling STEPO to perform one (forward) timestep. The tendencies are then supplied to MO3DPRJ which computes the required increments, using the 'explicit' (Machenhauer) or the 'implicit' scheme (or both, after which the results are merged). The increments are added to the original spectral fields and the process is iterated NITNMI (by default 2) times.

7.5 COMPUTATION OF J_c BASED ON NMI

In the notation of (7.1), the penalty term J_c is defined by

$$J_c = \varepsilon \| (\dot{\mathbf{x}} - \dot{\mathbf{x}}_b)_G \|^2 \tag{7.2}$$

where ε is an empirically chosen weighting factor, $\mathbf{x}$ is the current state of the control variable and $\mathbf{x}_b$ is the background. The norm $\| \ \|^2$ is based on a weighted sum of squares of spectral coefficients. Only the first NVMOD vertical modes are included in the evaluation of (7.2).

J_c is computed by calling the routine COSJC. Control passes through JCCOMP to NMIJCTL, where J_c is evaluated by calling STEPO twice, then projecting the differences in the tendencies on to the gravity modes via MO3DPRJ, and finally computing J_c in NMICOST.

7.6 DIGITAL FILTER INITIALIZATION

Digital filter initialization consists in removing high frequency oscillations from the temporal signal represented by the meteorological fields. A general description of digital filter initialization can be found in Lynch (1993). It can be implemented as a strong constraint by filtering the model fields at the beginning of each forecast or as a weak constraint as described in Gustafsson (1992) and Gauthier and Thépaut (2001).

Time oscillations exceeding a cut-off frequency $\omega_c = (2\pi)/T_c$ can be filtered by applying a digital filter to a time series $f_k = f(t_k)$ for $f_k = k\Delta t$, Δt being the timestep. This proceeds by doing a convolution of $f(t)$ with a step function $h(t)$ so that

$$f \bullet h(t_N) = \sum_{k=-\infty}^{\infty} h_k f_{N-k}$$

The step function h_k is found to be

$$h_k = \frac{\sin(\omega_c k \Delta t)}{k\pi}$$

In practice, the convolution is restricted to a finite time interval of time span T_s. We can write $T_s = 2M\Delta t$ and

$$f \bullet h(t_0) = \sum_{k=-M}^{M} \alpha_k f_k$$

with $\alpha_k = -h_{-k}$. This truncation introduces Gibbs oscillations which can be attenuated by introducing a Lanczos window which implies that the weights α_k are defined as $\alpha_k = -h_{-k}W_k$ with

$$W_k = \frac{\sin((k\pi)/(M+1))}{(k\pi)/(M+1)}$$

An alternative which is used at ECMWF has been proposed by Lynch (1997) to use a Dolph–Chebyshev window in which case

$$W_k = \frac{1}{2M+1}\left[1 + 2r\sum_{m=0}^{M} T_{2M}(x_0 \cos\theta_m/2)\cos m\theta_k\right] \tag{7.3}$$

where $1/x_0 = \cos(\pi\Delta t)/\tau_s$, $1/r = \cosh(2M \operatorname{acosh} x_0)$, $\theta_k = (k2\pi)/M$ and T_{2M} is the Chebyshev polynomial of degree $2M$. The time span of the window is chosen so that $\tau_s = M\Delta t$.

7.7 IMPLEMENTATION OF DFI AS A WEAK CONSTRAINT IN 4D-VAR

In the context of 4D-Var data assimilation, the digital filter is used as a weak constraint. A penalty term is added to the cost function and replaces the NMI based penalty term. The implementation is based

on Gauthier and Thépaut (2001). The filtered increments are calculated as

$$\delta\bar{\mathbf{X}}(t_{N/2}) = \sum_{k=0}^{N} \alpha_k \delta\mathbf{X}(t_k)$$

where N is the number of time steps in the minimisation and delta represents increments.

During each integration of the tangent linear model in the inner loop of the 4D-Var, the digital filter is applied to the increments and the partial sum accumulated by EDIGFIL. This gives a filtered increment valid at the mid-point of the assimilation window (arrays RACCSPA2 and RACCSPA3). The value of the non-filtered increment valid at the same time is also stored by ECOPSP in arrays RSTOSPA2 and RSTOSPA3. These routines are called in CNT4TL each time-step.

The adjoint model integration calculates the gradient in EDIGFILAD. The gradient is obtained by a single backward integration of the adjoint model. The adjoint calculations associated with the digital filter is a virtually cost free addition to the adjoint observation cost function calculations.

The filtering weights used by the digital filter are calculated during the setup phase by SUEFW. The default is to use a Dolph–Chebyshev non-recursive filter (NTPDFI = 4, see equation (7.3))

The weak constraint term which is added to the cost function is the moist energy norm of the departure between those two states times a weight factor. All these computations are conducted in spectral space and applied to the spectral fields. The default mode at ECMWF is to redefine the norm so the digital filter is only applied to divergence (LDIVONLY = .TRUE.). A larger weight factor (ALPHAG = 100.) is in that case used in the weak constraint term.

The norm of the departure is computed in two steps. In EVJCDFI, the difference between RACCSPA2/ RACCSPA3 and RSTOSPA2/RSTOSPA3 is computed and multiplied with ALPHAG for the subset of wave numbers and vertical levels associated with each processor. The cost contribution is calculated for each wavenumber and vertical level using the specified norm. The contributions for all wavenumbers, levels and variables are gathered on each processor by GATHERCOST2 and the total cost contribution is summed in array RSUMJCDFI for each variable and level. Finally, in EVCOST, the contributions from each variable and level are added to obtain the value of the penalty term.

Chapter 8
Diagnostics

Table of contents

8.1 INTRODUCTION

8.1.1 Influence matrix diagnostic in 4D-Var

The influence matrix is used in ordinary least-squares applications for monitoring statistical multiple-regression analyses. Concepts related to the influence matrix provide diagnostics on the influence of individual data on the analysis, the analysis change that would occur by leaving one observation out, and the effective information content (degrees of freedom for signal) in any sub-set of the analysed data. The corresponding concepts have been derived in the context of linear statistical data assimilation in numerical weather prediction. An approximate method to compute the diagonal elements of the influence matrix (the self-sensitivities) has been developed for a large-dimension variational data assimilation system (the 4D-Var system of ECMWF).

8.1.2 How to compute self-sensitivity

From clear-case branch: in /scripts/gen/ifsmin, namelist NAMVAR for the forecast error calculation (fce) must have LANOBS = .TRUE..

In directory /vol/verify/sms/userID/exp/include, namelist NAMVAR in ifsmin (for forecast error calculation) must have LANOBS = .TRUE..

Self-sensitivities are stored in ODB (CCMA) in the word *an_sens_obs*. The final self-sensitivities are computed by dividing the content in *an_sens_obs* to (the content of) *final_obs_error* for every observation assimilated.

A Fortran program is available (the program has been developed by Carla Cardinali, work is in progress to provide a user-version) to unload self-sensitivity from ODB and visual it with Metview.

8.2 OBSERVATIONAL INFLUENCE AND SELF-SENSITIVITY FOR A DA SCHEME

8.2.1 Linear statistical estimation in numerical weather prediction

Data assimilation systems for NWP provide estimates of the atmospheric state $\mathbf{x}$ by combining meteorological observations $\mathbf{y}$ with prior (or background) information $\mathbf{x}_b$. A simple Bayesian Normal model provides the solution as the posterior expectation for $\mathbf{x}$, given $\mathbf{y}$ and $\mathbf{x}_b$. The same solution can be achieved from a classical *frequentist* approach, based on a statistical linear analysis scheme providing the best linear unbiased estimate (Talagrand, 1997) of $\mathbf{x}$, given $\mathbf{y}$ and $\mathbf{x}_b$. The optimal GLS solution to

the analysis problem (see Lorenc, 1986) can be written

$$\mathbf{x}_\mathrm{a} = \mathbf{K}\mathbf{y} + (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{x}_\mathrm{b} \tag{8.1}$$

The vector $\mathbf{x}_\mathrm{a}$ is the 'analysis'. The gain matrix $\mathbf{K}(n \times p)$ takes into account the respective accuracies of the background vector $\mathbf{x}_\mathrm{b}$ and the observation vector $\mathbf{y}$ as defined by the $n \times n$ covariance matrix $\mathbf{B}$ and the $p \times p$ covariance matrix $\mathbf{R}$, with

$$\mathbf{K} = (\mathbf{R}^{-1} + \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1} \tag{8.2}$$

Here, $\mathbf{H}$ is a $p \times n$ matrix interpolating the background fields to the observation locations, and transforming the model variables to observed quantities (e.g. radiative transfer calculations transforming the models temperature, humidity and ozone into brightness temperatures as observed by several satellite instruments). In the 4D-Var context introduced below, $\mathbf{H}$ is defined to include also the propagation in time of the atmospheric state vector to the observation times using a forecast model.

Substituting (8.2) into (8.1) and projecting the analysis estimate onto the observation space, the estimate becomes

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{x}_\mathrm{a} = \mathbf{H}\mathbf{K}\mathbf{y} + (\mathbf{I}_p - \mathbf{H}\mathbf{K})\mathbf{H}\mathbf{x}_\mathrm{b} \tag{8.3}$$

It can be seen that the analysis state in observation space ($\mathbf{H}\mathbf{x}_\mathrm{a}$) is defined as a sum of the background (in observation space, $\mathbf{H}\mathbf{x}_\mathrm{b}$) and the observations $\mathbf{y}$, weighted by the $p \times p$ square matrices $\mathbf{I} - \mathbf{H}\mathbf{K}$ and $\mathbf{H}\mathbf{K}$, respectively.

In this case, for each unknown component of $\mathbf{H}\mathbf{x}$, there are two data values: a real and a 'pseudo' observation. The additional term in (8.3) includes these pseudo-observations, representing prior knowledge provided by the observation-space background $\mathbf{H}\mathbf{x}_\mathrm{b}$. From (8.3), the analysis sensitivity with respect to the observations is obtained

$$\mathbf{S} = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}} = \mathbf{K}^\mathrm{T}\mathbf{H}^\mathrm{T} \tag{8.4}$$

Similarly, the analysis sensitivity with respect to the background (in observation space) is given by

$$\frac{\partial \hat{y}}{\partial(\mathbf{H}\mathbf{x}_\mathrm{b})} = \mathbf{I} - \mathbf{K}^\mathrm{T}\mathbf{H}^\mathrm{T} = \mathbf{I}_p - \mathbf{S} \tag{8.5}$$

We focus here on the expressions (8.4) and (8.5). The influence matrix for the weighted regression DA scheme is actually more complex, but it obscures the dichotomy of the sensitivities between data and model in observation space.

The (projected) background influence is complementary to the observation influence. For example, if the self-sensitivity with respect to the ith observation is S_{ii}, the sensitivity with respect the background projected at the same variable, location and time will be simply $1 - \mathrm{S}_{ii}$. It also follows that the complementary trace, $\mathrm{tr}(\mathbf{I} - \mathbf{S}) = p - \mathrm{tr}(\mathbf{S})$, is not the *df* for noise but for background, instead. That is the weight given to prior information, to be compared to the observational weight $\mathrm{tr}(\mathbf{S})$. These are the main differences with respect to standard LS regression. Note that the different observations can have different units, so that the units of the cross-sensitivities are the corresponding unit ratios. Self-sensitivities, however, are pure numbers (no units) as in standard regression. Finally, as long as $\mathbf{R}$ is diagonal, (2.6) is assured, but for more general non-diagonal $\mathbf{R}$-matrices it is easy to find counter-examples to that property.

Inserting (8.1) into (8.4), we obtain

$$\mathbf{S} = \mathbf{R}^{-1}\mathbf{H}(\mathbf{B}^{-1} + \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\mathrm{T} \tag{8.6}$$

As $(\mathbf{B}^{-1} + \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H})$ is equal to the analysis error covariance matrix $\mathbf{A}$, we can also write $\mathbf{S} = \mathbf{R}^{-1}\mathbf{H}\mathbf{A}\mathbf{H}^\mathrm{T}$.

8.2.2 Approximate calculation of self-sensitivity in a large variational analysis system

In a optimal variational analysis scheme, the analysis error covariance matrix $\mathbf{A}$ is approximately the inverse of the matrix of second derivatives (the Hessian) of the cost function J, i.e. $\mathbf{A} = (\mathbf{J})^{-1}$ (Rabier and Courtier, 1992). Given the large dimension of the matrices involved, $\mathbf{J}$ and its inverse cannot be computed explicitly. Following Fisher and Courtier (1995) we use an approximate representation of the Hessian based on a truncated eigen-vector expansion with vectors obtained through the Lanczos algorithm. The calculations are performed in terms of a transformed variable P, $\mathbf{P} = \mathbf{L}^{-1}(\mathbf{x} - \mathbf{x}_\mathrm{b})$, with $\mathbf{L}$ chosen such that $\mathbf{B}^{-1} = \mathbf{L}^\mathrm{T}\mathbf{L}$. The transformation $\mathbf{L}$ thus reduces the covariance of the prior to the identity matrix. In variational assimilation $\mathbf{L}$ is referred to as the change-of-variable operator (Courtier *et al.*, 1998).

$$\mathbf{J}''^{-1} \simeq \mathbf{B} - \sum_{i=1}^{M} \frac{1-\lambda_i}{\lambda_i}(\mathbf{L}v_i)(\mathbf{L}v_i)^T \tag{8.7}$$

The summation in (8.7) approximates the variance reduction **B-A** due to the use of observations in the analysis. $(\lambda_i, \mathrm{v}_i)$ are the eigen-pairs of $\mathbf{A}$. In ECMWF's operational data assimilation system, the variances of analysis error are computed according to this method. The variances are inflated to provide estimates of short-term forecast (background) error variances to be used as background errors in the next analysis cycle (Fisher, 1996). The Hessian eigen-vectors are also used to precondition the minimization (Fisher and Andersson, 2001). The computed eigen-values are not used to minimize the cost function but only to estimate the analysis covariance matrix. It is well known, otherwise, that the minimization algorithm is analogous to the conjugate-gradient algorithm. Because the minimum is found within an iterative method, the operational number of iterations is sufficient to find the solution (with required accuracy) but does not provide a sufficient number of eigen-pairs to estimate the analysis-error variances.

The diagonal of the background error covariance matrix $\mathbf{B}$ in (8.7) is also computed approximately, using the randomisation method proposed by Fisher and Courtier (1995). From a sample of N random vectors $\mathbf{u}_\mathrm{i}$ (in the space of the control-vector), drawn from a population with zero mean and unit Gaussian variance, a low-rank representation of $\mathbf{B}$ (in terms of the atmospheric state variables $\mathbf{x}$) is obtained by

$$\mathbf{B} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{L}u_i)(\mathbf{L}u_i)^T \tag{8.8}$$

This approximate representation of $\mathbf{B}$ has previously been used by Andersson *et al.* (2000) to diagnose background errors in terms of observable quantities, i.e. $\mathbf{HBH}^\mathrm{T}$.

Inserting (8.7) and (8.8) into (8.6) we arrive at an approximate method for calculating $\mathbf{S}$, that is practical for a large dimension variational assimilation (both 3D and 4D-Var). This is given by

$$\mathbf{S} = \mathbf{R}^{-1}\mathbf{H}\left[\frac{1}{N}\sum_{i=1}^{N}(\mathbf{L}u_i)(\mathbf{L}u_i)^T + \sum_{i=1}^{M}\frac{1-\lambda_i}{\lambda_i}(\mathbf{L}v_i)(\mathbf{L}v_i)^T\right]\mathbf{H}^T \tag{8.9}$$

Only the diagonal elements of $\mathbf{S}$ are computed and stored in ODB – that is, the analysis sensitivities with respect to the observations, or self-sensitivities S_{ii}. The cross-sensitivity S_{ij} for $i \neq j$, that represents the influence of the jth observation to the analysis at the ith location, is not computed. Note that the approximation of the first term is unbiased, whereas the second term is truncated such that variances are underestimated. For small M the approximate S_{ii} will tend to be over-estimates. For the extreme case $M = 0$, (8.9) gives $\mathbf{S} = \mathbf{R}^{-1}\mathbf{HBH}^\mathrm{T}$ which in particular can have diagonal elements larger than one if elements of $\mathbf{HBH}^\mathrm{T}$ are larger than the corresponding elements of $\mathbf{R}$. The number of Hessian vectors operationally computed is $M = 40$ and the number of random $\mathbf{B}$ vectors is $N = 50$.

In general, in the operational system, 15% of the global influence is due to the assimilated observations in any one analysis, and the complementary 85% is the influence of the prior (background) information, a short-range forecast containing information from earlier assimilated observations. About 25% of the observational information is currently provided by surface-based observing systems, and 75% by satellite systems.

Low-influence data points usually occur in data-rich areas, while high-influence data points are in data-sparse areas or in dynamically active regions. Background error correlations also play an important role: High correlation diminishes the observation influence and amplifies the importance of the surrounding real and pseudo observations (prior information in observation space). Incorrect specifications of background and observation error covariance matrices can be identified, interpreted and better understood by the use of influence matrix diagnostics for the variety of observation types and observed variables used in the data assimilation system.

Self-sensitivities cannot be larger than one (they are bounded in the interval zero to one) but, because of the small number of eigenpair we can compute, S_{ii} can be greater than one. Approximations in both of the two terms of (8.9) contribute to the problem. In the second term the number of Hessian eigen-vectors is truncated to M. The term is therefore underestimated, and S_{ii} will tend to be over-estimated. The degree of over-estimation depends on the structure of the covariance reduction matrix **B-A**.

For an analysis in which observations lead to strongly localised covariance reduction (such as the humidity analysis with its short co-variance length scales ~180 km, and large observational impacts) a large M is required to approximate **B-A** accurately. The approximate computation is mostly affecting the self-sensitivities close to the upper bound leaving the self-sensitivities <0.7 almost unaffected.

To conclude, the self-sensitivity provides a quantitative measure of the observation influence in the analysis. In robust regression, it is expected that the data have similar self-sensitivity (sometimes called leverage) – that is, they exert similar influence in estimating the regression line. Disproportionate data influence on the regression estimate can have different reasons: First, there is the inevitable occurrence of incorrect data. Second, influential data points may be legitimately occurring extreme observations. However, even if such data often contain valuable information, it is constructive to determine to which extent the estimate depends on these data. Moreover, diagnostics may reveal other patterns e.g. that the estimates are based primarily on a specific sub-set of the data rather than on the majority of the data. In the context of 4D-Var there are many components that together determine the influence given to any one particular observation. First there is the specified observation error covariance **R**, which is usually well known and obtained simply from tabulated values. Second, there is the background error covariance **B**, which is specified in terms of transformed variables that are most suitable to describe a large proportion of the actual background error covariance. The implied covariance in terms of the observable quantities is not immediately available for inspection, but it determines the analysis weight given to the data. Third, the dynamics and the physics of the forecast model propagate the covariance in time, and modify it according to local error growth in the prediction. The influence is further modulated by data density. Low influence data points occur in data-rich areas while high influence data points are in data-sparse regions or in dynamically active areas. Background error correlations also play an important role. In fact, very high correlations drastically lessen the observation influence in favour of background influence and amplify the influence of the surrounding observations.

With the approximate method used here, out-of-bound self-sensitivities occur if the Hessian representation based on an eigen-vector expansion is truncated, especially when few eigen-vectors are used. However, this problem affects only a small percentage of the self-sensitivities computed in this study, and in particular those that are closer to one. Remaining values greater than one can be due to large background to observation error ratio, which is one factor that is known to contribute towards ill-conditioning and poor convergence of the 4D-Var algorithm.

8.3 THE GRADIENT TEST

If LTEST = .TRUE. a gradient test will be performed both before and after minimization. This is done by the routine GRTEST. In the gradient test a test value t_1 is computed as the ratio between a perturbation of the co-t-function and its first-order Taylor expansion using

$$t_1 = \lim_{\delta\chi \to 0} \frac{J(\chi + \delta\chi) - J(\chi)}{\langle \nabla J, \delta\chi \rangle} \tag{8.10}$$

with $\delta\chi = -\alpha\nabla J$. Repeatedly increasing α by one order of magnitude, printing t_1 at each step should show t_1 approaching one, by one order of magnitude at a time, provided $J(\chi)$ is approximately quadratic

over the interval $[\chi, \chi + \delta\chi]$. The near linear increase in the number of 9's in the print of t_1 over a wide range of α (initially as well as after minimization) proves that the coded adjoint is the proper adjoint for the linearization around the given state χ.

The behaviour of the cost function in the vicinity of χ in the direction of the gradient ∇J is also diagnosed by several additional quantities for each α. The results are printed out on lines in the log-file starting with the string 'GRTEST:'. To test the continuity of J, for example, a test value t_0 is computed with

$$t_0 = \frac{J(\chi + \delta\chi)}{J(\chi)} - 1 \tag{8.11}$$

and printed. For explanation of other printed quantities see the routine GRTEST itself. A range of additional test was introduced in Cy28r2.

Chapter 9
Observation processing

Table of contents

9.1 BASIC PRINCIPLES

The ECMWF Data Assimilation Observation Processing System prior to Cy26r1 (April 2003) was roughly split in two parts.

(i) Non-IFS observation processing modules.
(ii) IFS integrated observation processing module.

Originally, the main difference in function between these two parts was based on whether information about a field (e.g. first guess) was required or not. Thus, the observation processing functions for which field information was not required were dealt with by the non-IFS modules, whereas the IFS itself dealt with those observation processing functions for which field information was needed.

The non-IFS observation processing came in two main parts.

(i) Preparation and massaging of input BUFR data.
(ii) Creation of two data structure; one acceptable by the IFS as input and the other one for archiving purposes (observation feedback).

The first part of the non-IFS observation processing, which is still intact, consists of a number of modules: PRE1CRAD (further split by instrument type), PREOBS, PREOBS_WAVE, PREGEOS, PREREO3 and PRESCAT. Without going into too many details here, the main theme for all of them is to prepare input BUFR data in an appropriate form for further processing. This also involves performing preliminary data thinning. As such, this part is preserved even after the major change which occurred with Cy26r1.

The second part of the non-IFS observation processing consisted of two modules: OBSPROC and OBSORT. The main task of OBSPROC was to prepare input BUFR data in a form to be used by the analysis, whereas OBSORT dealt with any issues related to parallel computing. In this context OBSORT was not doing anything on its own; it was normally called by OBSPROC to ensure efficient parallelisation. During an analysis cycle OBSPROC is executed twice: just before and just after the IFS. The task before the IFS, called MAKECMA or for short MKCMA, performed a number of observation processing functions.

(i) Read in and crack input BUFR data.
(ii) Carry out preliminary data checks.
(iii) Perform necessary variable changes.
(iv) Assign observation errors.
(v) Create CMA data structure recognised by the IFS.
(vi) Etc.

On the other hand the task of OBSPROC just after the IFS, called FEEDBACK, was to create BUFR feedback. This was done by appending the input BUFR data with analysis-related information (departures, flags, events, etc.).

Cy26r1 saw a major revision in this area. Observation processing modules OBSPROC and OBSORT, as well as the CMA observation data structure, have been phased out. Hence, MKCMA and FEEDBACK tasks as we knew them were made obsolete. However, a new data structure, the ODB, as well as two new observation processing modules (BUFRTOODB and ODBTOBUFR) have been introduced. Most of the observation processing functions earlier performed by the MKCMA task within OBSPROC have now been included in the IFS. It is only purely BUFR related processing functions that have now been taken over by BUFRTOODB and ODBTOBUFR.

- BUFRTOODB, together with MERGEODB, runs just before the IFS and is called MAKEODB. Effectively what it does is to read input BUFR data and create initial ODB which is formally acceptable by the IFS.
- ODBTOBUFR together with MATCHUP runs just after the IFS and is called ODB2BUFR.

Both MAKEODB and ODB2BUFR have been developed and are handled by the Operations Department.

The OBSORT observation processing functions have now almost entirely been incorporated into the ODB software.

As mentioned earlier most of the observation processing functions of OBSPROC are now integrated in the IFS. These newly integrated IFS observation processing functions are now known as "MAKE CMA REPLACEMENT" or for short MKCMARPL.

Here we will mostly concentrate on the IFS integrated observation processing whereas the other parts of the ECMWF documentation will deal with the remaining aspects of observation processing.

9.2 MAIN MKCMARPL TASKS AND FUNCTIONS

9.2.1 Basic observation processing setup

In order to perform the observation processing functions, a number of basic observation processing setups are carried out at the very beginning of initialising the IFS. This is done by calling several routines in addition to all other routines needed to setup the IFS (see Fig. 9.1).

- Program MASTER calls CNT0 which in turn calls SU0YOMA.
- SU0YOMA calls (among other routines) SUOAF from which SUCMOCTP, SUEVENTS, SUCODES, SUFLTXT and SUCMA are called. SUCMOCTP defines the ODB observation types and code types, and SUEVENTS, SUCODES and SUFLTXT define analysis events, various codes used and flags naming conventions.
- SUCMA calls SUCMAF which then calls several subroutines: SUCMAD1, SUCMAD2, SUCMAHFP, SUCMAHOP, SUCMBDFP and SUCMBDTP. These routines define the structure of ODB Data Descriptor Records (DDRs) as well as the ODB packing patterns (bit structure) employed for header and body respectively.

9.2.2 Invoking, initializing and controlling the MKCMARPL

The MKCMARPL run is initiated by the MKCMARPL subroutine (see Fig. 9.2). This routine is only invoked in the SCREENING run of the IFS. It is called, together with some of its additional setup routines via subroutine SUOBS. The additional setup routines called at this level are: SUANCT, DEFRUN, SULIM, SULEVLAY, SUSATRET, SUVNMB, SUSCRE0, SUOBSORT, SETCOM, DEPERERR, SUERRORS, INIERSCA and INISSMIP.

- MKCMARPL is namelist driven and in DEFRUN a logical variable LMKCMARPL is defined. By default LMKCMARPL = .TRUE. but it can be overwritten via namelist NAMOBS. Furthermore, many other parameters and switches are defined in DEFRUN and some of them can also be overwritten via namelists.
- SUANCT and SULIM define some additional analysis constants and limits.
- SULEVLAY and SUSATRET define analysis related level/layer and satellite retrieval parameters, respectively.
- SUVNMB declares variable numbers.
- SUSCRE0, SUOBSORT and SETCOM define flag limits, identify ambiguous moving platforms, initialise observation sorting, and provide some general observation common variables.
- DEPERERR and SUERRORS deal with observation error statistics definitions. SUERRORS calls SUPERERR to define observation persistence errors and SUOBSERR to define prescribed observation errors.
- INIERSCA and INISSMIP deal with initialising SCATT and SSMI processing.

The next step is to find out if it is a SCREENING run and if so to check if it is a MKCMARPL run as well. In the case of a MKCMARPL run all aspects of the observation processing before the screening are dealt with by calling MKCMARPL (more about it in Subsection 9.2.3). After MKCMARPL has finished there are several ways to proceed. These depend on the status of LMKCMARPLO and LRPLSWAPOUT

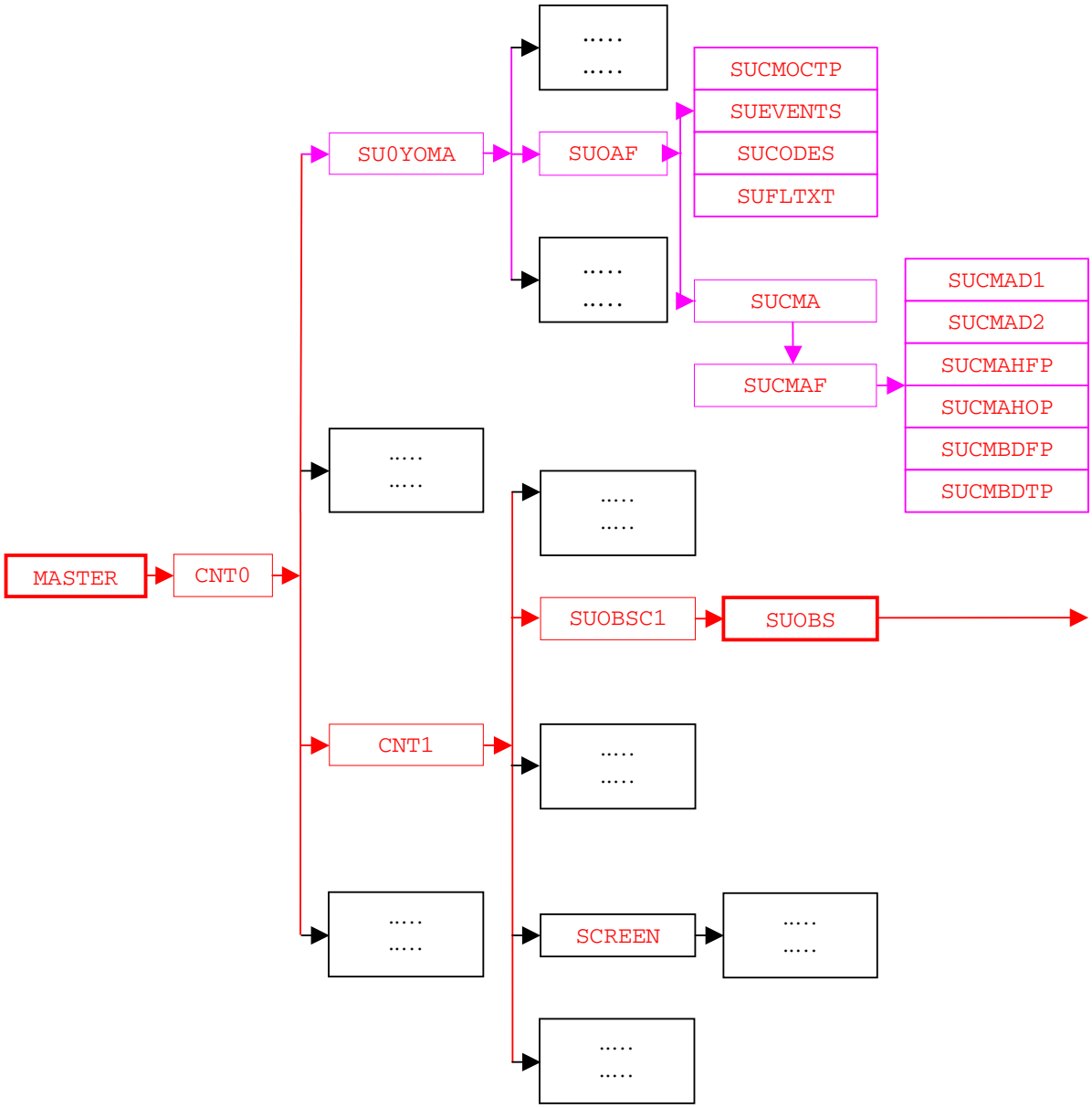

Figure 9.1 *Simplified IFS observation pre-processing flow diagram (MASTER). Colour coding scheme: (a) routines in red boxes perform observation pre-processing, (b) routines in pink boxes carry out observation pre-processing set up, and (c) routines in black boxes are not directly involved in observation pre-processing. Fig. 9.2 continues the flow diagram from SUOBS.*

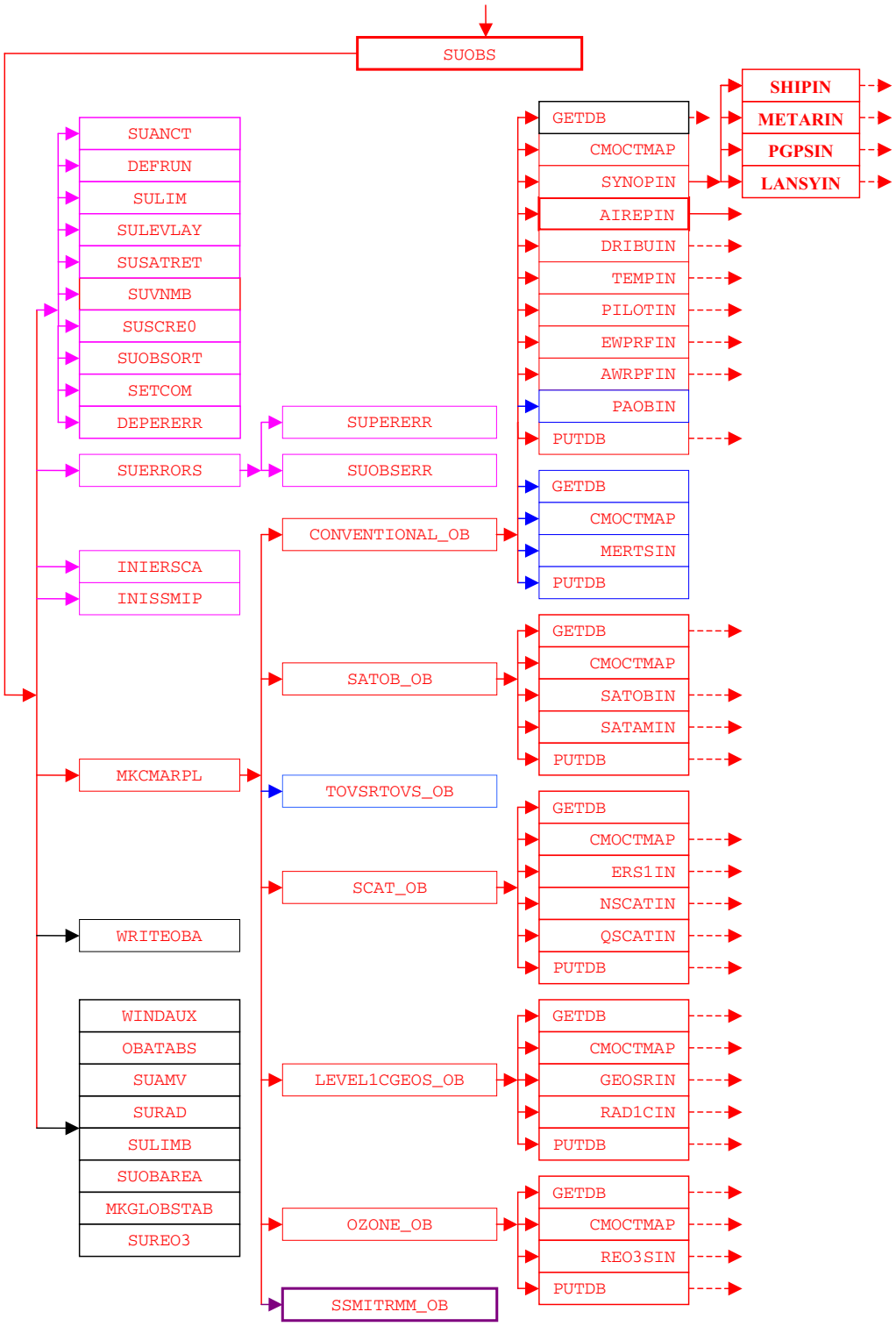

Figure 9.2 (Continued from Fig. 9.1) *Simplified IFS observation pre-processing flow diagram (SUOBS). Colour coding scheme: (a) routines in red boxes perform observation pre-processing, (b) routines in pink boxes carry out observation pre-processing set up, (c) routines in black boxes are not directly involved in observation pre-processing, (d) routines in blue boxes are obsolete and (e) routines in plum boxes are awaiting revision. Fig. 9.3 continues the flow diagram from AIREPIN.*

logical switches (NAMOBS namelist). If LRPLSWAPOUT = .TRUE. the ODB is swapped out and if LMKCMARPLO = .TRUE. the ODB is written out and the run terminated. Both of these options are not normally used and their use is for diagnostics/debugging purposes. Once the MKCMARPL work has been completed the remainder of SUOBS will execute as before. Thus, calls to WRITEOBA, WINDAUX, OBATABS, SUAMV, SURAD, SULIMB, SUOBAREA, MKGLOBSTAB and SUREO3 are issued.

In the context of operational running, the MKCMARPL related switches are set:

LMKCMARPL = .TRUE. LRPLSWAPOUT = .FALSE. LMKCAMRPLO = .FALSE.

9.2.3 MKCMARPL

The main purpose of MKCMARPL is to control the IFS observation pre-processing. Observation pre-processing at this stage is done in groups of observations. At the moment there are seven groups: CONVENTIONAL, SATOB, TOVS/RTOVS, SCATT, LEVEL1C/GEOSS, OZONE and SSMI/TRMM observations. For each group a separate subroutine is called: CONVENTIONAL_OB, SATOB_OB, TOVSRTOVS_OB, SCAT_OB, LEVEL1CGOES_OB, OZONE_OB and SSMITRMM_OB. These routines are just cover or hat routines for the actual work to be carried out underneath. However, TOVSRTOVS_OB and SSMITRMM_OB are currently not called because TOVSRTOVS_OB is obsolete and SSMITRMM_OB is waiting for a major revision.

Each cover routine would call the ODB to get the observations it wants to process. This is done by calling the ODB GETDB subroutine. As the observations are brought, in one or more worker routines would be called to perform the observation processing functions. Once the worker routines have finished the control is handed back to the cover routine. The next step in the cover routine is to return observations back to the ODB database. This is done by calling the ODB PUTDB routine. In some of these cover routines several calls to GETDB/PUTDB might be issued. This is because there may be sufficient differences between similar data to justify a slightly different approach in their pre-processing. For example under the CONVENTIONAL_OB routine there are two calls to a GETDB and PUTDB pair. The first call deals with all conventional observations except SATEMs; the second call deals with the SATEMs. As indicated earlier, between each GETDB and PUTDB a number of observations type or code type designed worker routines are called.

- CONVENTIONAL_OB calls the following worker routines: SYNOPIN, AIREPIN, DRIBUIN, TEMPIN, PILOTIN, EWPRFIN, AWPRFIN, PAOBIN and MERTSIN. A worker routine name indicates which observations it is dealing with.
- SATOB_OB calls SATOBIN and SATAMIN.
- SCAT_OB calls ERS1IN, NSCATIN and QSCATIN.
- LEVEL1CGEOS_OB calls RAD1CIN and GOESRIN.
- OZONE_OB calls only REO3SIN.

9.2.4 Basic observation handling routines

The observation pre-processing worker routines referred to in Subsection 9.2.3, names of which always end with “IN”, are the basic observation handling routines. They all follow more or less the same logic. As an example consider AIREPIN which deals with AIREP observations (see Fig. 9.3).

The first thing which is done is to define the instrument specification (OBINSTP) followed by preliminary quality control check both at the report level (PRLMCHK) as well as at the data level (GETSETE and AIREPBE).

- PRLMCHK calls REPSEL and TIMDIF to do report selection according to preset criteria and to find out time difference between analysis time and the actual observation time, respectively.
- GETSETE makes a local copy of a given observation variable and its related parameters from an ODB supplied array.
- After updating the local copy, AIREPBE is called to return the updated local copy back to the ODB supplied array.

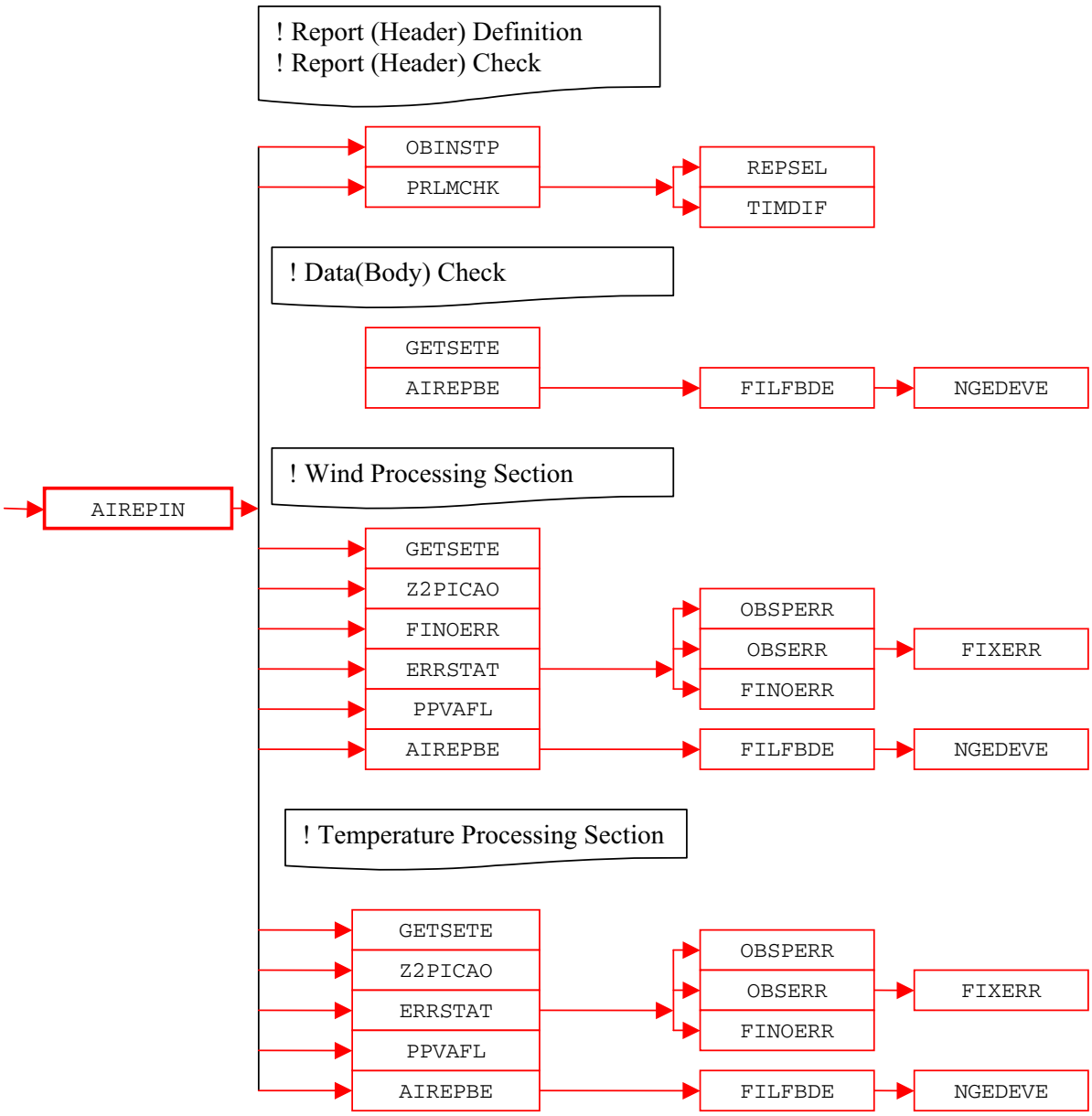

Figure 9.3 (Continued from Fig. 9.2) *Simplified IFS observation pre-processing flow diagram (AIREPIN). Colour coding scheme: routines in red boxes perform observation pre-processing.*

The preliminary quality control at the report level consists of making sure that observation position, date and time are reasonable. Furthermore, as there is a possibility of excluding certain observations via the NAMOBS namelist, a check is made of whether the observation is actually wanted at this stage. Once the report level check is passed attention is turned to the data itself. Each datum is checked against predefined list of expected data. If not in the list, datum is rejected and a warning message issued. At this stage it is also ensured that missing indicators used are unique.

After the preliminary phase attention is turned to getting data in the right form and shape for further usage. Thus, in the case of an AIREP observation, this is done in sections of available variables: wind and temperature.

(i) **Wind.** There are four wind variables: wind direction (*DDD*), wind force (*FFF*), u and v components. For each of these variables the first thing which is done is to get a local copy of it together with its related parameters from an ODB supplied array (GETSETE). Once a variable is made available

locally a check is made to ensure that the vertical coordinate is pressure; if instead of pressure a flight level is supplied it is converted into pressure by assuming a standard ICAO atmosphere (Z2PICAO). If the variable in question is either u or v, then DDD and FFF are converted into u and v wind components. Furthermore, for each of the four variables appropriate observation error statistics are assigned (ERRSTAT, FINOERR). Also, if any flags are set at this stage an appropriate word in the local copy is updated (PPVAFL). Finally, an updated local copy of an observed quantity and its related parameters are returned back into the ODB (AIREPBE).

(ii) **Temperature.** In the case of temperature only one observed variable is dealt with. The pattern of making a local copy (GETSETE), ensuring that pressure is the vertical coordinate (Z2PICAO), assigning the observation error statistics (ERRSTAT), updating flags (PPVAFL) and returning an updated local copy back to the ODB (AIREPBE) is repeated.

As just mentioned ERRSTAT deals with assigning observation errors for a given observation variable. ERRSTAT first calls OBSPERR to assign observation persistence error; then it calls OBSERR which in turn calls FIXERR to assign prescribed observation error. It is worth mentioning that observation errors themselves are already predefined at an earlier stage (SUERRORS).

The pattern of activities outlined for AIREPIN is repeated more or less in the other worker routines. However, the SYNOPIN routine is first split further into SHIPIN, METATRIN, PGPSIN and LANSYIN. This is because SHIP, METAR, GPS and SYNOP LAND observations are sufficiently different to justify a separate worker routine. Furthermore, LANSYIN is somewhat more complicated than AIREPIN. One of the reasons for this is that we have to distinguish between low and high level stations.

9.3 OBSERVATION TYPES, SUBTYPES AND CODE TYPES

All observations, both in the BUFR and ODB contexts, are split into a number of observation types. The observation types are then further divided into observation code types (ODB) and observation subtypes (BUFR). Although BUFR observation types and subtypes are not directly used in the IFS they are defined here. BUFR observation types and subtypes are mapped into ODB observation types and code types before the IFS (i.e. the MERGEODB step).

9.3.1 BUFR observation types and subtypes

There are eight BUFR observation types. However, the number of subtypes differs between observation types; they are listed in Table 9.1.

9.3.2 ODB observation and code types

There are ten ODB observation types and, as with BUFR, there are a different number of code types for each of them. It is a reasonable to question why the BUFR and ODB observation types and sub or code types are different. The answer is a historic one. The ODB observation types and code types have been used before BUFR came in to existence and as an international code it was difficult to impose our practice on the others. Also, there was not enough enthusiasm on our side to switch to the BUFR ones. The ODB observation types and code types are listed in Table 9.2.

9.3.3 Mapping between ODB and BUFR observation types, code types and subtypes

As indicated in Subsection 9.3.2 the coexistence of different codes used for BUFR and ODB observation types and the subtype and code type requires a mapping from one to another. This is given in Table 9.3.

9.4 VARIABLES

Different quantities are observed by different observing systems. It is only a subset of observed quantities that are used in the analysis and most of them are used in their original form. However, some of them are transformed into the ones actually used by the analysis. This transformation, or a change of variable, may also include retrieval from satellite data if they are independent from the background model fields. The original variables may be kept with the derived ones so that first guess departures can be assigned

Table 9.1 *BUFR observation types and subtypes.*

Observation Type		Subtype	
Code	**Name**	**Code**	**Name**
0	Land Surface	1	Land SYNOP
		3	Automatic Land SYNOP
		9	Abbreviated Land SYNOP
		110	GPS
		140	METAR
1	Sea Surface	9	SHIP
		11	SHIP
		13	Automatic SHIP
		19	Reduced SHIP
		21	DRIBU
		22	BATHY
2	Upper Air Sounding	91	Land PILOT
		92	SHIP PILOT
		95	Wind Profiler (American)
		96	Wind Profiler (European/Japanese)
		101	Land TEMP
		102	SHIP TEMPS
		103	DROP TEMP
		104	ROCOB
		105	SHIP ROCOB
		106	Mobile TEMP
3	Satellite Sounding	0	High Resolution TOVS
		51	High Resolution TOVS
		53	RTOVS
		54	ATOVS
		55	ATOVS
		57	ATOVS
		61	Low Level Temperature SATEM
		62	High Level SATEM
		63	PWC SATEM
		65	Merged SATEM
		71	Low Level TOVS
		72	High Level TOVS
		73	PWC TOVS
		75	Merged TOVS
		129	TRMM
		130	TMI
		161	PAOB
		206	OZONE Retrieved Layers
4	AIREP	142	AIREP
		143	COLBA
		144	AMDAR
		145	ACARS
5	SATOB	82	Temperature and Wind
		83	Wind Only
		84	Temperature only
		85	Temperature only
		86	High Resolution VIS Wind

Table 9.1 *Continued.*

Observation Type		Subtype	
Code	**Name**	**Code**	**Name**
		87	AMV
		89	Geostationary Clear Sky Radiances (GRAD)
		189	Geostationary Clear Sky Radiances (GRAD)
12	ERS/SSMI	8	ERS 1
		122	ERS 1
		127	SSMI
		136	NSCAT
		137	
253	PAOB	161	PAOB

for both. Furthermore, if an observed variable is transformed then, if necessary, so is its observation error statistics. Also, in the case of an off-time SYNOP observation, the observed surface pressure may be adjusted.

9.4.1 Observed variables

The exact list of what is observed or present in the list of BUFR observation types and sub types (Table 9.3) is long. Therefore Table 9.4 just lists (per observation types) those variables which are of interest at present.

9.4.2 Derived variables

Variables which are transformed for further use by the analysis are as follows.

(i) Wind direction (DDD) and force (FFF) are transformed into wind components (u and v) for SYNOP, AIREP, SATOB, DRIBU, TEMP and PILOT observations.
(ii) Temperature (T) and dew point (Td) are transformed into relative humidity (RH) for SYNOP and TEMP observations, with a further transformation of the RH into specific humidity (Q) for TEMP observations.
(iii) SCATTEROMETER backscatters (σ^0's) are transformed into a pair of ambiguous wind components (u and v); this actually involves a retrieval according to some model function describing the relationship between winds and σ^0's and requires a fair bit of computational work.
(iv) Mean layer temperature is transformed into thickness (DZ) for SATEM and TOVS observations.

All these variable transformations, except for the σ^0's transformation, are more or less trivial ones. The wind components are worked out as

$$u = -FFF \sin\left(DDD\frac{\pi}{180}\right)$$
$$v = -FFF \cos\left(DDD\frac{\pi}{180}\right)$$

The RH is derived using

$$RH = \frac{F(Td)}{F(T)}$$

where function F of either T or Td is expressed as

$$F(T) = a\frac{R_{\text{dry}}}{R_{\text{vap}}}e^{b\frac{T-T_0}{T-c}}$$

Table 9.2 *ODB observation types and code types.*

Observation Type		Code Type	
Code	**Name**	**Code**	**Name**
1	SYNOP	11	Land SYNOP
		14	Automatic Land SYNOP
		16	French RADOME
		21	SHIP
		22	Abbreviated SHIP
		23	SHRED
		24	Automatic SHIP
		140	METAR
		110	GPS
2	AIREP	41	CODAR
		141	AIREP
		142	Simulated AIREP
		144	AMDAR
		145	ACARS
		241	COLBA
3	SATOB	88	SATOB
		89	High Resolution VIS wind
		90	AMV
		188	SST
4	DRIBU	63	BATHY
		64	TESAC
		160	ERS as DRIBU
		165	DRIBU
5	TEMP	35	Land TEMP
		36	SHIP TEMP
		37	Mobile TEMP
		39	ROCOB
		40	SHIP ROCOB
		135	DROP TEMP
		137	Simulated TEMP
6	PILOT	32	Land PILOT
		33	SHIP PILOT
		34	American Wind Profiler
		131	Japanese Wind Profiler
		132	Mobile Wind Profiler
		134	European Wind Profiler
7	SATEM	86	GTS SATEM
		184	High Resolution Simulated TOVS
		185	High Resolution Simulated DWL SATEM
		186	High Resolution SATEM
		200	GTS BUFR SATEM 250km
		201	GTS BUFR Clear Radiances
		202	GTS BUFR Retrieved Profiles/Clear Radiances
		210	ATOVS/GRAD
		211	RTOVS
		212	TOVS
		215	SSMI

Table 9.2 *Continued.*

Observation Type		Code Type	
Code	**Name**	**Code**	**Name**
8	PAOB	164	PAOB
9	SCATTEROMETER	8	SCATTEROMETER
		122	SCATTEROMETER
		210	SCATTEROMETER
10	RAW RADIANCE		

where $T_0 = 273.16$ K, $a = 611.21$, $b = 17.502$, $c = 32.19$, $R_{\text{dry}} = 287.0597$ and $R_{\text{vap}} = 461.5250$ are constants.

The specific humidity Q is worked out by using

$$Q = RH \frac{A}{1 - RH\left(\frac{R_{\text{vap}}}{R_{\text{dry}}} - 1\right)A}$$

with function A is expressed as

$$A = \min\left[0.5, \frac{F(T)}{P}\right]$$

where P is pressure. Q is assigned in the RH2Q subroutine.

Scatterometer wind retrieval is dealt with in Section 9.6.

9.4.3 Adjusted variables

The only observed quantity which is adjusted is the SYNOP's surface pressure (P_{s}). This is done by using pressure tendency (P_{t}) information, which in turn may be first adjusted. P_{t} is adjusted only in the case of SYNOP SHIP data for the ship movement.

The ship movement information is available from input data in terms of ship speed and direction, which are first converted into ship movement components U_{s} and V_{s}. The next step is to find pressure gradient ($\partial p/\partial x$ and $\partial p/\partial y$) given by

$$\frac{\partial p}{\partial x} = C(A_1 u - A_2 v)\frac{1}{2}$$
$$\frac{\partial p}{\partial y} = -C(A_1 u + A_2 v)$$

where u and v are observed wind components, and $A_1 = 0.94$ and $A_2 = 0.34$ are the sine and cosine of the angle between the actual and geostrophic winds. C is the Coriolis term multiplied by a drag coefficient (D) so that

$$C = 2\Omega D \sin\theta$$

where, θ is the latitude and $\Omega = 0.7292 \times 10^{-4} s^{-1}$ is the angular velocity of the earth and D is expressed as

$$D = GZ$$

$G = 1.25$ is an assumed ratio between geostrophic and surface wind over sea and $Z = 0.11$ kgm^{-3} is an assumed air density. Now the adjusted pressure tendency (P_{t}^{a}) is found as

$$P_{\text{t}}^{\text{a}} = P_{\text{t}} - \left(U_{\text{s}}\frac{\partial p}{\partial x} + V_{\text{s}}\frac{\partial p}{\partial y}\right)$$

Finally, the adjusted surface pressure (P_{s}^{a}) is found as

$$P_{\text{s}}^{\text{a}} = P_{\text{s}} - P_{\text{t}}^{\text{a}}\Delta t$$

where, Δt is a time difference between analysis and observation time. Of course in the case of non-SHIP data $P_{\text{t}}^{\text{a}} \equiv P_{\text{t}}$. Subroutine PTENDCOR is used for this adjustment.

Table 9.3 *Mapping between ODB and BUFR observation types, code types and subtypes.*

ODB (Observation Type, Code Type)		BUFR (Observation Type, Subtype)
ODB(1, 11)	↔	BUFR[(0,1);(0,9)]
ODB(1, 14)	↔	BUFR(0,3)
ODB(1, 21)	↔	BUFR(1,9)
ODB(1, 22)	↔	BUFR(?,?)
ODB(1, 23)	↔	BUFR(1,19)
ODB(1, 24)	↔	BUFR(1,13)
ODB(1,110)	↔	BUFR(0,110)
ODB(1,140)	↔	BUFR(0,140)
ODB(2,41)	↔	BUFR(?,?)
ODB(2,141)	↔	BUFR(4,142)
ODB(2,142)	↔	BUFR(?,?)
ODB(2,144)	↔	BUFR(4,144)
ODB(2,145)	↔	BUFR(4,145)
ODB(2,241)	↔	BUFR(4,143)
ODB(3,88)	↔	BUFR[(5,82);(5,83);(5,84);(5,85)]
ODB(3,89)	↔	BUFR(5,86)
ODB(3,90)	↔	BUFR(5,87)
ODB(3,188)	↔	BUFR(?,?)
ODB(4,63)	↔	BUFR(1,23)
ODB(4,64)	↔	BUFR(1,22)
ODB(4,160)	↔	BUFR(?,?)
ODB(4,165)	↔	BUFR(1,21)
ODB(5,35)	↔	BUFR(2,101)
ODB(5,36)	↔	BUFR(2,102)
ODB(5,37)	↔	BUFR(2,106)
ODB(5,39)	↔	BUFR(2,104)
ODB(5,40)	↔	BUFR(2,105)
ODB(5,135)	↔	BUFR(2,103)
ODB(5,137)	↔	BUFR(?,?)
ODB(6,32)	↔	BUFR(2,91)
ODB(6,33)	↔	BUFR(2,92)
ODB(6,34)	↔	BUFR(2,94)
ODB(6,131)	↔	BUFR(2,95)
ODB(6,134)	↔	BUFR(2,95)
ODB(7,86)	↔	BUFR[(3,61);(3,62);(3,63);(3,65)]
ODB(7,184)	↔	BUFR(?,?)
ODB(7,185)	↔	BUFR(?,?)
ODB(7,186)	↔	BUFR[(3,71);(3,72);(3,73);(3,75)]
ODB(7,200)	↔	BUFR(?,?)
ODB(7,201)	↔	BUFR(?,?)
ODB(7,202)	↔	BUFR(?,?)
ODB(7,206)	↔	BUFR(?,?)
ODB(7,210)	↔	BUFR[(3,54);(5,89)]
ODB(7,211)	↔	BUFR(3,53)
ODB(7,212)	↔	BUFR[(3,0);(3,51)]
ODB(7,215)	↔	BUFR(12,127)
ODB(8,180)	↔	BUFR(253,154)
ODB(9,8)	↔	BUFR(12,8)
ODB(9,122)	↔	BUFR(12,122)
ODB(9,300)	↔	BUFR(12,136)
ODB(9,301)	↔	BUFR(12,137)
ODB(9,511)	↔	BUFR(?,?)
ODB(10,1)	↔	BUFR(?,?)

Table 9.4 *Observed variables.*

Observation Type		
BUFR	**ODB**	**Observed Variable**
Land Surface	Land SYNOP	Surface Pressure (P_s) 10 m Wind Direction/Force (DDD/FFF) 2 m Temperature (T_{2m}) 2 m Dew Point (Td_{2m}) Pressure Tendency (P_t) Cloud Information Precipitation Information Snow Depth (Sd) Etc.
Sea Surface	SHIP SYNOP, DRIBU	Surface Pressure (P_s) 10 m Wind Direction/Force (DDD/FFF) 2 m Temperature (T_{2m}) 2 m Dew Point (Td_{2m}) Etc.
Upper Air Sounding	TEMP, PILOT	10m/Upper Air Wind Direction/Force (DDD/FFF) 2m/Upper Air Temperature (T_{2m}/T) 2m/Upper Air Dew Point (Td_{2m}/Td) Geopotential Height (Z) Etc.
Satellite Sounding	SATEM	Mean Layer Temperature Precipitable Water Content (PWC) Brightness Temperature (Tb)
AIREP	AIREP	Upper Air Wind Direction/Force (DDD/FFF) Temperature (T)
SATOB	SATOB	Upper Air Wind Direction/Force (DDD/FFF) Brightness Temperature (Tb)
ERS/SSMI	SCATTEROMETER	Backscatter (σ^0) Brightness Temperature (Tb)

9.4.4 Codes for variables

To provide easy recognition of 'observed' variables each of them is assigned a numerical code. These codes are then embedded in ODB reports. There are 76 codes used so far. These codes are defined in subroutine SUVNMB. For the sake of completeness these codes are listed in Table 9.5.

9.5 OBSERVATION ERROR STATISTICS

Three types of observation errors are dealt with at the observation pre-processing level.

(i) Persistence observation error.
(ii) Prescribed observation error.
(iii) Combination of the two above called the final observation error.

9.5.1 Persistence observation error

The persistence error is formulated in such a way to reflect its dependence on the following.

(i) Season.

Table 9.5 *Numbering of variables in the ODB.*

No.	Code	Name	Unit
1	3	Wind Component (u)	ms^{-1}
2	4	Wind Component (v)	ms^{-1}
3	1	Geopotential (Z)	m^2s^{-2}
4	57	Thickness (DZ)	m^2s^{-2}
5	29	Relative Humidity (RH)	numeric
6	9	Precipitable Water Content (PWC)	kgm^{-2}
7	58	2 m Relative Humidity (RH_{2m})	numeric
8	2	Temperature	K
9	59	Dew Point	K
10	39	2 m Temperature (T_{2m})	K
11	40	2 m Dew Point (Td_{2m})	K
12	11	Surface Temperature (Ts)	K
13	30	Pressure Tendency (P_t)	Pa/3h
14	60	Past Weather (W)	WMO Code 4561
15	61	Present Weather (WW)	WMO Code 4677
16	62	Visibility (V)	WMO Code 4300
17	63	Type of High Clouds (C_H)	WMO Code 0509
18	64	Type of Middle Clouds (C_M)	WMO Code 0515
19	65	Type of Low Clouds (C_L)	WMO Code 0513
20	66	Cloud Base Height (N_h)	m
21	67	Low Cloud Amount (N)	WMO Code 2700
22	68	Additional Cloud Group Height ($h_s h_s$)	m
23	69	Additional Cloud Group Type (C)	WMO Code 0500
24	70	Additional Cloud Group Amount (N_s)	WMO Code 2700
25	71	Snow Depth (Sd)	m
26	72	State of Ground (E)	WMO Code 0901
27	73	Ground Temperature ($T_g T_g$)	K
28	74	Special Phenomena ($S_p S_p$)	WMO Code 3778
29	75	Special Phenomena ($s_p s_p$)	WMO Code 3778
30	76	Ice Code Type (R_s)	WMO Code 3551
31	77	Ice Thickness ($E_s E_s$)	WMO Code 1751
32	78	Ice (I_s)	WMO Code 1751
33	79	Time Period of Rain Information ($t_r t_r$)	hour
34	80	6 Hour Rain Amount	kgm^{-2}
35	81	Maximum Temperature (JJ)	K
36	82	Ship Speed (V_s)	ms^{-1}
37	83	Ship Direction (D_s)	degree
38	84	Wave Height ($H_w H_w$)	m
39	85	Wave Period ($P_w P_w$)	s
40	86	Wave Direction ($D_w D_w$)	degree
41	87	General Cloud Group	WMO Code
42	88	Relative Humidity from Low Clouds	numeric
43	89	Relative Humidity from Middle Clouds	numeric
44	90	Relative Humidity from High Clouds	numeric
45	91	Total Amount of Clouds	WMO Code 20011
46	92	6 Hour Snowfall	m
47	110	Surface Pressure (P_s)	Pa
48	111	Wind Direction	degree
49	112	Wind Force	ms^{-1}
50	119	Brightness Temperature (Tb)	K
51	120	Raw Radiance	K

Table 9.5 *Continued.*

No.	Code	Name	Unit
52	121	Cloud Amount from Satellite	%
53	122	Backscatter (σ^0)	dB
54	5	Wind Shear ($\partial u/\partial z$)	s^{-1}
55	6	Wind Shear ($\partial v/\partial z$)	s^{-1}
56	41	u_{10m}	ms^{-1}
57	42	v_{10m}	ms^{-1}
58	19	Layer Relative Humidity	numeric
59	200	Auxiliary Variable	numeric
60	123	Cloud Liquid Water (Q_l)	$kgkg^{-1}$
61	124	Ambiguous v	ms^{-1}
62	125	Ambiguous u	ms^{-1}
63	7	Specific Humidity (Q)	$kgkg^{-1}$
64	126	Ambiguous Wind Direction	degree
65	127	Ambiguous Wind Speed	ms^{-1}
66	8	Vertical Speed	ms^{-1}
67	56	Virtual Temperature (Tv)	K
68	206	Ozone	Dobson
69	156	Height	m
70	215	SSM/I Pseudo Variable	kgm^{-2}
71	160	Past Weather	numeric
72	130	Pressure Tendency Characteristics	numeric
73	12	Sea Water Temperature	K
74	192	Radar Reflectivity	Db
75	128	Atmospheric Path Delay in Satellite Signal	m
76	162	Radio Occultation Bending Angle	Rad

(ii) Actual geographical position of an observation.

Seasonal dependency is introduced by identifying three regimes.

(i) Winter hemisphere.
(ii) Summer hemisphere.
(iii) Tropics.

The positional dependency is then introduced to reflect the dependence on the precise latitude within these three regimes.

The persistence error calculation is split into two parts. In the first part the above dependencies are expressed in terms of factors a and b which are defined as

$$a = \sin\left(2\pi\frac{d}{365.25} + \frac{\pi}{2}\right)$$

and

$$b = 1.5 + a\left\{0.5\min\left[\frac{\max(\theta, 20)}{20}\right]\right\}$$

where d is a day of year and θ is latitude.

The persistence error for time difference between analysis and observation Δt is then expressed as a function of b with a further dependence on latitude and a maximum persistence error E_{maxpers} for 24 hour given by

$$E_{\text{pers}} = \frac{E_{\text{maxpers}}}{6}[1 + 2\sin(|2\theta|b\Delta t)]$$

Table 9.6 *Observation persistence errors of maximum 24-hour wind* (u, v)*, height* (Z) *and temperature* (T)*.*

Variable (unit)	1000–700 hPa	699–250 hPa	249–0 hPa
u, v (ms^{-1})	6.4	12.7	19.1
Z (m)	48	60	72
T (K)	6	7	8

where Δt is expressed as a fraction of a day. The E_{maxpers} have the values shown in Table 9.6.

Subroutine SUPERERR is used to define all relevant points in order to carry out this calculation, and is called only once during the general system initialization. The calculation of the actual persistence error is dealt with by OBSPERR.

9.5.2 Prescribed observation errors

Prescribed observational errors have been derived by statistical evaluation of the performance of the observing systems, as components of the assimilation system, over a long period of operational use. The prescribed observational errors are given in the Tables 9.7, 9.8 and 9.9. Currently, observational errors are defined for each observation type that carries the following quantities.

(i) Wind components.
(ii) Height.
(iii) Temperature.
(iv) Humidity.

As can be seen from the tables of prescribed observation errors, they are defined at standard pressure levels but the ones used are interpolated to the observed pressures. The interpolation is such that the observation error is kept constant below the lowest and above the highest levels, whereas in between it is interpolated linearly in $\ln p$. Several subroutines are used for working out the prescribed observation error: SUOBSERR, OBSERR, FIXERR, THIOERR and PWCOERR.

- SUOBSERR defines observation errors for standard pressure levels.
- OBSERR and FIXERR calculate the actual values.
- THIOERR and PWCOERR are two specialised subroutines to deal with thickness and PWC errors.

Relative humidity observation error RH_{err} is either prescribed or modelled. More will be said about the modelled RH_{err} in Subsection 9.5.3. RH_{err} is prescribed only for TEMP and SYNOP data. RH_{err} is preset to 0.17 for TEMP and 0.13 for SYNOP. However, if $RH < 0.2$ it is increased to 0.23 and to 0.28 if $T < 233$ K for both TEMP and SYNOP.

9.5.3 Derived observation errors

Relative humidity observation error, RH_{err}, can also be expressed as function of temperature T so that

$$RH_{\mathrm{err}} = \min[0.18, \min(0.06, -0.0015T + 0.54)]$$

This option is currently used for assigning RH_{err}.

Specific humidity observation error, Q_{err}, is a function of RH, RH_{err}, P, P_{err}, T and T_{err}, and formally can be expressed as

$$Q_{\mathrm{err}} = Q_{\mathrm{err}}(RH, RH_{\mathrm{err}}, P, P_{\mathrm{err}}, T, T_{\mathrm{err}})$$

or

$$Q_{\mathrm{err}} = RH_{\mathrm{err}} F_1(RH, P, T) + \frac{RH\ P_{\mathrm{err}}}{P} F_2(RH, P, T) + RH\ T_{\mathrm{err}}(RH, P, T)$$

Table 9.7 *Prescribed RMS observation errors for the u and v wind components* (ms^{-1}).

R RAD All	SCATT All	PAOB All	SATEM All	PILOT All	TEMP All	DRIBU All	SATOB All	AIREP	AIREP All but AIREP	SYNOP All	Observation Type Code	Type
n/a	2.00	n/a	n/a	1.80	1.80	1.80	2.00	2.86	2.46	3.00	1000	Levels (hPa)
n/a	2.00	n/a	n/a	1.80	1.80	1.80	2.00	2.91	2.51	3.00	850	
n/a	2.00	n/a	n/a	1.90	1.90	1.80	2.00	2.96	2.56	3.00	700	
n/a	2.00	n/a	n/a	2.10	2.10	1.80	3.50	3.11	2.71	3.40	500	
n/a	2.00	n/a	n/a	2.50	2.50	1.80	4.30	3.21	2.81	3.60	400	
n/a	2.00	n/a	n/a	2.60	2.60	1.80	5.00	3.26	2.86	3.80	300	
n/a	2.00	n/a	n/a	2.50	2.50	1.80	5.00	3.31	2.91	3.20	250	
n/a	2.00	n/a	n/a	2.50	2.50	1.80	5.00	3.36	2.96	3.20	200	
n/a	2.00	n/a	n/a	2.40	2.40	1.80	5.00	3.31	2.91	2.40	150	
n/a	2.00	n/a	n/a	2.20	2.20	1.80	5.00	3.16	2.76	2.20	100	
n/a	2.00	n/a	n/a	2.10	2.10	1.80	5.00	3.06	2.66	2.00	70	
n/a	2.00	n/a	n/a	2.00	2.00	1.80	5.00	3.06	2.66	2.00	50	
n/a	2.00	n/a	n/a	2.10	2.10	1.80	5.00	3.26	2.86	2.00	30	
n/a	2.00	n/a	n/a	2.30	2.30	1.80	5.00	3.46	3.06	2.50	20	
n/a	2.00	n/a	n/a	3.00	3.00	1.80	5.70	3.76	3.36	3.00	10	

Table 9.8 *Prescribed RMS height observation errors* (m).

R RAD All	SCATT All	PAOB All	SATEM All	PILOT All	TEMP All	DRIBU All	SATOB All	AIREP All	SYNOP Automatic SHIP	SYNOP Manual SHIP	SYNOP Automatic Land	SYNOP Manual Land	Observation Type Code	Observation Type
n/a	n/a	24.0	n/a	4.3	4.3	4.97	n/a	n/a	4.2	7.1	4.2	5.6	1000	Levels (hPa)
n/a	n/a	24.0	n/a	4.4	4.4	4.97	n/a	n/a	4.2	7.1	5.4	7.2	850	
n/a	n/a	24.0	n/a	5.2	5.2	4.97	n/a	n/a	4.2	7.1	6.45	8.6	700	
n/a	n/a	24.0	n/a	8.4	8.4	4.97	n/a	n/a	4.2	7.1	9.07	12.1	500	
n/a	n/a	24.0	n/a	9.8	9.8	4.97	n/a	n/a	4.2	7.1	11.17	14.9	400	
n/a	n/a	24.0	n/a	10.7	10.7	4.97	n/a	n/a	4.2	7.1	14.1	18.8	300	
n/a	n/a	24.0	n/a	11.8	11.8	4.97	n/a	n/a	4.2	7.1	19.05	25.4	250	
n/a	n/a	24.0	n/a	13.2	13.2	4.97	n/a	n/a	4.2	7.1	20.77	27.7	200	
n/a	n/a	24.0	n/a	15.2	15.2	4.97	n/a	n/a	4.2	7.1	24.3	32.4	150	
n/a	n/a	24.0	n/a	18.1	18.1	4.97	n/a	n/a	4.2	7.1	29.55	39.4	100	
n/a	n/a	24.0	n/a	19.5	19.5	4.97	n/a	n/a	4.2	7.1	37.72	50.3	70	
n/a	n/a	24.0	n/a	22.5	22.5	4.97	n/a	n/a	4.2	7.1	44.47	59.3	50	
n/a	n/a	24.0	n/a	25.0	25.0	4.97	n/a	n/a	4.2	7.1	52.35	69.8	30	
n/a	n/a	24.0	n/a	32.0	32.0	4.97	n/a	n/a	4.2	7.1	72.0	96.0	20	
n/a	n/a	24.0	n/a	40.0	40.0	4.97	n/a	n/a	4.2	7.1	85.65	114.2	10	

Table 9.9 *Prescribed RMS temperature observations error* (K).

R RAD All	SCATT All	PAOB All	SATEM All	PILOT All	TEMP All	DRIBU All	SATOB All	AIREP AIREP	AIREP All but AIREP	SYNOP SHIP	SYNOP Land	Observation Type Code	Observation Type
n/a	n/a	n/a	n/a	n/a	1.40	1.8	n/a	1.40	1.65	1.8	2.0	1000	Levels (hPa)
n/a	n/a	n/a	n/a	n/a	1.25	1.5	n/a	1.18	1.43	1.8	1.5	850	
n/a	n/a	n/a	n/a	n/a	1.10	1.3	n/a	1.00	1.25	1.8	1.3	700	
n/a	n/a	n/a	n/a	n/a	0.95	1.2	n/a	0.98	1.23	1.8	1.2	500	
n/a	n/a	n/a	n/a	n/a	0.90	1.3	n/a	0.96	1.21	1.8	1.3	400	
n/a	n/a	n/a	n/a	n/a	1.00	1.5	n/a	0.05	1.20	1.8	1.5	300	
n/a	n/a	n/a	n/a	n/a	1.15	1.8	n/a	0.95	1.20	1.8	1.8	250	
n/a	n/a	n/a	n/a	n/a	1.20	1.8	n/a	1.06	1.31	1.8	1.8	200	
n/a	n/a	n/a	n/a	n/a	1.25	1.9	n/a	1.18	1.43	1.8	1.9	150	
n/a	n/a	n/a	n/a	n/a	1.30	2.0	n/a	1.30	1.55	1.8	2.0	100	
n/a	n/a	n/a	n/a	n/a	1.40	2.2	n/a	1.40	1.65	1.8	2.2	70	
n/a	n/a	n/a	n/a	n/a	1.40	2.4	n/a	1.50	1.75	1.8	2.4	50	
n/a	n/a	n/a	n/a	n/a	1.40	2.5	n/a	1.60	1.85	1.8	2.5	30	
n/a	n/a	n/a	n/a	n/a	1.50	2.5	n/a	1.80	2.05	1.8	2.5	20	
n/a	n/a	n/a	n/a	n/a	2.10	2.5	n/a	2.10	2.35	1.8	2.5	10	

where functions F_1, F_2 and F_3 are given by

$$F_1(RH,T,P) = \frac{A}{\left[1 - RH\left(\frac{R_{\rm vap}}{R_{\rm dry}} - 1\right)A\right]^2}$$

$$F_2(RH,T,P) = \frac{\left\{\left[1 - RH\left(\frac{R_{\rm vap}}{R_{\rm dry}} - 1\right)A\right] + \left(\frac{R_{\rm vap}}{R_{\rm dry}} - 1\right)A\right\}}{\left[1 - RH\left(\frac{R_{\rm vap}}{R_{\rm dry}} - 1\right)A\right]^2}$$

$$F_3(RH,T,P) = \frac{Ab(T_0 - c)}{(T-c)^2}\left\{\left[1 - \left(\frac{R_{\rm vap}}{R_{\rm dry}} - 1\right)A\right] RH\, A\left(\frac{R_{\rm vap}}{R_{\rm dry}} - 1\right)\right\}$$

At present only the first term of the above expression for $Q_{\rm err}$ is taken into account (dependency on relative humidity). Subroutine RH2Q is used to evaluate $Q_{\rm err}$.

Surface pressure observation error $Ps_{\rm err}$ is derived by multiplying the height observation error $Z_{\rm err}$ by a constant:

$$Ps_{\rm err} = 1.225\, Z_{\rm err}$$

However, the $Ps_{\rm err}$ may be reduced if the pressure tendency correction is applied. For non-SHIP data the reduction factor is 4, whereas for SHIP data the reduction factor is either 2 or 4, depending on if the $P_{\rm t}$ is adjusted for SHIP movement or not.

The thickness observation error ($DZ_{\rm err}$) is derived from $Z_{\rm err}$.

9.5.4 Final (combined) observation error

In addition to the prescribed observation and persistence errors, the so called final observation error is assigned at this stage too. This is simply a combination of observation and persistence errors given by

$$F_{\rm OE} = \sqrt{O_{\rm E}^2 + P_{\rm E}^2}$$

where $F_{\rm OE}, O_{\rm E}$ and $P_{\rm E}$ are final, prescribed and persistence observation errors, respectively. The subroutine used for this purpose is FINOERR.

9.6 SCATTEROMETER DATA

9.6.1 Overview

The processing for three different scatterometer instruments, each having its own geometrical and geophysical characteristic has been coded in IFS. These are scatterometer data from the AMI instrument on-board the ERS-1 and ERS-2 platforms, the NSCAT instrument on-board ADEOS-I, and the SeaWinds instrument on-board QuikSCAT.

As for a number of other observation types, scatterometer data need to be transformed into the variables used by the analysis within the IFS. This transformation converts the backscatter measurements acquired by the instrument (triplets for ERS and quadruplets for NSCAT and QuikSCAT) into the two ambiguous u and v wind components that will actually be assimilated into the IFS. The (empirical) relation between wind and backscatter is described by a geophysical model function (GMF).

At the implementation of Cy28r1 (March 2004), data from ERS-2 (re-introduced after an interruption since January 2001, and using a new GMF), and from QuikSCAT (introduced in Cy24r3) were actively assimilated.

9.6.2 ERS data

For ERS-1 and ERS-2 scatterometer data the inversion task is performed within the MKCMARPL run by the scatterometer data handling subroutine SCAT_OB, especially by its core subroutine ERS1IF (ERS-1 interface). Like SCAT_OB, ERS1IF deals only with one scatterometer report at a time.

The main purpose of ERS1IF is to retrieve the wind components by inverting the geophysical model function. Some quality control is also done in the process, based on the quality information provided with the raw data, as well as on the residual from the wind retrieval, reflecting the agreement between the measurements and their theoretical wind dependency. The procedure closely follows the PRESCAT wind retrieval and ambiguity removal scheme developed for the ERS-1 scatterometer (Stoffelen and Anderson, 1997), though the original geophysical model function CMOD4 has been replaced by CMOD5.

Backscatter data from the ERS-1 and ERS-2 platform are provided into wind-vector cells (WVC) with a spatial resolution of 25 km. Swath width is 500 km, defining 19 cells in across-track direction. In this direction, in IFS, ERS data is thinned by only selecting predefined cells (subroutine SCAQC). The reason for this are known across-track variations in data quality. Along-track, this is not an issue, and therefore, thinning is performed according to the standard procedure (THINN) as also used for other satellite data. Both SCAQC and THINN result in an average selection of cells determined by the parameter NTHINSCA; a default value of NTHINSCA = 4 results in a selection of every fourth cell giving a thinning resolution of 100 km.

Moreover, bias corrections are applied, both in terms of backscatter and wind speed, particularly to compensate for any change in the instrumental calibration and to ensure consistency between the retrieved and model winds.

9.6.3 QuikSCAT data

For data from the SeaWinds scatterometer on-board the QuikSCAT satellite, the task of wind inversion is performed at an earlier stage of the processing. The program performing this task, QSCAT25TO50KM, is a part of a scatterometer library called SCAT.

Backscatter data from the QuikSCAT platform is also provided at a resolution of 25 km. Across the 1,800 wide swath 76 cells are defined. Rather than data thinning, in SCAT, a 50 km product is created instead which contains information about the backscattering from the four underlying original sub-cells. For the 38 across-track cells defined in this way the outer 4 at either side of the swath are, due to their known reduced quality, not assimilated. The weight of the scatterometer cost function (defined in routine HJO) of each 50 km wind vector cell is reduced by a factor four, which effectively mimics the assimilation of a 100 km product.

9.6.4 NSCAT data

Due to its short lifetime of nine months, NSCAT data has never been part of the operational assimilation chain at ECMWF. Although, in IFS, NSCAT data can be handled, assimilation experiments are only possible after certain offline pre-processing of the data. The Research Department should be contacted for further information.

9.6.5 Wind retrieval

In general, the wind retrieval is performed by minimizing the distance between observed backscatter values σ^0_{o} and modelled backscatter values $\sigma^0_{\mathrm{m}i}$ given by

$$D(u) = \sum_{i}^{n} \frac{[(\sigma^0_{\mathrm{o}i})^p - \sigma^0_{\mathrm{m}i}(u)^p]^2}{k_p \left[\sum_j^n \sigma^0_{\mathrm{m}j}(u)^p\right]^2}$$

For ERS data, the sum is over triplets, while for QuikSCAT the sum may extend to 16 values (four 25 km sub-cells with each four observations). The quantity p is equal to unity for NSCAT and QuikSCAT. For ERS data, a value of $p = 0.625$ was introduced because it makes the underlying GMF more harmonic, which helps to avoid direction-trapping effects (Stoffelen and Anderson, 1997). The noise to signal ratio k_p provides an estimate for the relative accuracy of the observations.

The simulation of σ^0_m is for ERS-2 data based on the CMOD5 model function (replacing CMOD4 used for earlier cycles; although for ERA-40 experiments CMOD4 is still selected). For NSCAT data the NSCAT-2 GMF has been utilized. For QuikSCAT data, the choice of GMF is handled by a logical switch

LQTABLE in the SCAT library. By default LQTABLE = .TRUE. and the QSCAT-1 model function is used, otherwise, modelled backscatter values are based on the NSCAT-2 GMF. The minimization is achieved using a tabular form of the GMF, giving the value of the backscatter coefficient for wind speeds, direction and incidence angles discretized with, for ERS data, steps of 0.5 ms^{-1}, 5^0 and 1^0, respectively. For NSCAT and QuikSCAT data the corresponding values are 0.2 ms^{-1}, 2.5° and 1°. For ERS the table is read in the initialisation subroutine of the scatterometer observation processing INIERSCA. For QuikSCAT, this takes place in the QSCAT25TO50KM program in the PRESCAT task.

Up to four minima are kept at first. The first wind-vector solution is then defined as the minimum with the smallest residual, and the second one as the secondary minimum, etc.

The retrieval process is the subject of a few dedicated subroutines. For ERS, routine S0TOWIND directly returns two ambiguous wind vectors associated with a given triplet. This subroutine itself uses two specialized procedures, SPEEEST and MINIMA, to get a first guess estimate for the wind speeds yielding the lowest residuals and to search for the relative minima in the table, respectively. For QuikSCAT, a similar routine, INVERT50, returns up to four ambiguous wind vectors. The two corresponding routines it calls for first guess estimation and minimization are WSFG and MINIMA. In addition, it calls MEDIAS, a routine that calculates, based on the used data, the centre-of-gravity of the 50 km vector cell, and FFT99, a routine that was introduced to suppress numerical noise.

The retrieved wind components obviously play a major role in the definition of the scatterometer ODB reports. In addition, a logical parameter indicating whether no solution has actually been found is also transmitted for the subsequent processing applied in the IFS.

9.6.6 Quality control

Before calling for the wind retrieval, a first quality control step consists of checking the BUFR supplied instrumental quality flags set by ESA for ERS or JPL for QuikSCAT data. The data are processed only if they are complete and free from transmission errors or land, sea and/or ice contaminations.

For ERS it is also checked whether no arcings are present for any of the three antennae. Also for ERS, for each antenna k_p must stay smaller than 10%, and the missing packet number must be less than 10 to ensure that enough individual backscatter measurements have been averaged for estimating the value.

For QuikSCAT, it is checked whether the data are likely to be contaminated by rain. Since February 2000, the BUFR product provides a rain flag. This flag, which was developed by NASA/JPL, is based on a multidimensional histogram (MUDH) incorporating various quantities that may be used for the detection of rain (Huddleston and Stiles, 2000). Examples of such parameters are *mp_rain_probability* (an empirically determined estimate for the probability of a columnar rain rate larger than 2 m^2 hr^{-1}; typically values larger than 0.1 indicate rain contamination) and *nof_rain_index* (a rescaled normalized objective function – values larger than 20 give a proxy for rain). Since at the time of implementation, the quality of the JPL rain flag had not been fully confirmed, an alternative flag was established in house. Based on a study in which QuikSCAT winds were compared to collocated ECMWF first guess winds, a quality flag was introduced (tested in the REGROUP subroutine). This is given by

$$L_{\text{rain}} = \{nof\_rain\_index + 200\ mp\_rain\_probability > 30\}$$

Both *mp_rain_probability* and *nof_rain_index* are provided in the BUFR product (for details see Leidner *et al.*, 2000). When one of these quantities is missing, the above mentioned condition for the remaining quantity is used.

In addition, for QuikSCAT, it is verified whether inverted winds are well-defined, i.e. whether minima D(u) are sufficiently sharp. In practise this is mainly an issue for cells in the central part of the swath. Data is rejected when the angle between the most likely solution and its most anti-parallel one is less than 135° (routine SCAQC).

After wind inversion, a further check is then done on the backscatter residual associated to the rank-1 solution (or, more precisely, its square root called 'distance to the cone'). This quantity, representing the misfit between the observed and modelled backscatter values contains both the effects of instrumental noise and of GMF errors. Locally, these errors can become large when the measurements are affected by

geophysical parameters not taken into account by the GMF, such as sea-state or intense rainfall. For ERS the distance to the cone is normalized by its expected standard deviation, computed from the distance to the cone and an estimation of the geophysical noise as a function of wind speed and incidence angle. A triplet is considered rejected if the result exceeds a threshold of 3 (three-standard deviation test). For QuikSCAT data such a test is not performed.

Following these quality control checks, a flag is defined. This will be different from zero if any technical problem has been detected during the test of the ESA or JPL flag, or if either the distance to the cone has turned out to be too large (ERS) after wind retrieval or no solutions have even been found. This flag is used in the subsequent processing made in the screening, as described in the corresponding part of the IFS documentation.

In addition to a distance-to-cone test on single observations, a similar test is performed for averages for data within certain time slots. If these averages exceed certain values, all data within the considered time slot is suspected to be affected by an instrument anomaly, since geophysical fluctuations are expected to be averaged out when grouping together large numbers of data points. For ERS, node-wise averages are calculated for the default 4D-Var observation time slot (30 minutes since Cy24r3, 1 hour for older cycles) in the IFS routine SCAQC, and its rejection threshold (1.5 times average values) are defined in the IFS routine SUFGLIM. For QuikSCAT averages are considered over six-hourly data files and are calculated in the SCAT program DCONE_OC, using a threshold of 1.45 for any of WVC's 4 to 35.

9.6.7 ERS bias corrections

For ERS, two separate bias corrections are prepared in ERS1IF to improve the agreement of the retrieved winds with ECMWF model winds. First, a σ^0 bias correction is performed before the wind retrieval, by subtracting constant bias estimates from the raw backscatter measurements as a function of their antenna and WVC numbers. These bias estimates, derived from a routine comparison between the σ^0 measured by the scatterometer and the σ^0 simulated by CMOD5 (or CMOD4 for ERA-40) from the first guess winds of the ECMWF model, are supposed to account both for the variations that may occur in the instrumental calibration in time and for the residual defaults affecting the fit of the geophysical model function in the backscatter space.

A wind-speed bias correction is then added following the wind retrieval, in the form of a table that is dependent on retrieved wind speed and the measurement across-cell number. Its purpose is to match the scatterometer and model wind speeds over the entire wind-speed range so as to avoid introducing any speed-up or slow-down tendency in the assimilation process. Like the σ^0 bias correction, this wind-speed dependent bias correction relies on a direct comparison between scatterometer and model data, in which the wind speeds retrieved with the σ^0 bias correction are fitted as a function of those deduced from the model first guess according to a Maximum Likelihood Estimation (MLE) procedure.

The σ^0 wind-speed bias corrections are defined by two dedicated files read in the initialization subroutine INIERSCA, and containing appropriate coefficients for ERS-1 and ERS-2. Files are both model-cycle and date dependent. However, at the implementation of Cy28r1, the appropriate files had no effect for ERS-2 (i.e. containing only unity correction factors for σ^0, and zeros for wind bias corrections), since the CMOD5 geophysical model function had explicitly been tuned to result in, compared to ECMWF, unbiased winds.

It should be noted that the corrections made are not kept explicitly in the scatterometer ODB reports, where the main outputs are limited to the retrieved wind components as well as to the distances to the cone and the associated quality-control flags. Moreover, the original σ^0 measurements are also stored (and for ERS together with the ESA-retrieved wind speeds and directions), to allow subsequent data monitoring from the analysis feedback file.

9.6.8 QuikSCAT bias correction

For QuikSCAT data no bias corrections in σ^0 space is applied, though, wind-bias corrections are made. Such corrections are performed in three steps. First of all, wind speeds are reduced by 4% so that

$$v' = 0.96\, v$$

Where V is wind speed obtained from the INVERT50 subroutine. It was observed that the residual bias between QuikSCAT winds and ECMWF first guess winds depends on the value of *mp_rain_probability* (see Subsection 9.6.7). The motivation is that, for higher amounts of precipitation, a larger part of the total backscatter is induced by rain, leaving a smaller part for the wind signal. The correction applied is

$$v'' = v' - 20\langle\ mp\_rain\_probability\rangle$$

where $\langle\ \rangle$ denotes the average value over the 25 km sub-cells that were taken into account in the inversion (i.e. not over rain-flagged sub-cells). The maximum allowed correction is 2.5 ms^{-1}, which is seldom reached. Finally, for strong winds, QuikSCAT winds were found to be quite higher than their ECMWF first guess counterparts. In order to accommodate this, for winds stronger than 19 ms^{-1} the following correction is applied:

$$v''' = v'' - 0.2(v'' - 19.0)$$

The wind-speed bias corrections are applied in the QSCAT25TO50KM program of the SCAT library.

9.7 GEOSTATIONARY CLEAR-SKY RADIANCES OR BRIGHTNESS TEMPERATURES

9.7.1 Data, data producers and reception at ECMWF

Radiances from geostationary imagers of the Meteosat and the GOES series are used at ECMWF in the form of clear-sky radiances and corresponding brightness temperatures (CSR or CSBT, below referred to as CSR for brevity). The CSRs are area averages of those image pixels of a segment that have been diagnosed as clear-sky. This data pre-processing, including the cloud-detection, is carried by the satellite data providers.

Meteosat data are processed by EUMETSAT (Darmstadt, Germany). CSRs are produced for the water vapour and the infrared channel from hourly images for averaging segments of 16×16 pixels (about 80×80 km^2 areas at sub-satellite point). Data are encoded as BUFR and delivered via the GTS.

Data from the GOES satellites are processed by CIMSS/NESDIS (Madison, USA). CSRs are derived for all channels (visible, water vapour and infrared) and also produced hourly. GOES segments are 11×17 pixels (about 45×45 km^2 areas at sub-satellite point). Data are also BUFR encoded, but currently received at ECMWF via internet/ftp.

The content of the CSRs from Meteosat and GOES comprise clear-sky radiances for the channels indicated above as well as additional information such as location, time, satellite zenith angle, and fraction of clear and cloudy sky in the averaging area. For a complete list, see the data descriptors of the BUFR format. There are differences between the data provided by Meteosat and GOES, and changes to data format and content have occurred during the period for which CSR data have been received and treated. A common BUFR format has been approved (descriptors 301023 for imagers with up to 12 channels, 301024 for imagers with up to 3 channels). It is used by EUMETSAT since 2 December 2002; CIMSS should have provided GOES CSR data in this format from early 2003. Then also the standard deviation of the pixels within the CSR mean is provided as quality indication for all satellites.

After reception, data are recoded at ECMWF into a common BUFR format for storage in MARS and insertion into assimilation (IFS). For GOES data, some simple checks on reasonable time and location specifications have been included at this stage in order to trap erroneous data. In case of occurrence of any incorrect values, the whole data set (corresponding to one image) is rejected.

9.7.2 Overview over METEOSAT and GEOS imager CSR in the ECMWF archives

Table 9.10 gives a short summary of the CSR data stored at ECMWF either in MARS or in ECFS, including the BUFR subtype of the data. For more information on the actual content of the data see BUFR templates, bearing in mind that not all data items which can be encoded according to the CSR BUFR template are actually always provided (i.e. missing values). Incoming data from Meteosat and GOES are currently recoded into one BUFR format being the interface to observation processing and assimilation in IFS. This BUFR was originally designed for the Meteosat CSR. For the GOES data, not

Table 9.10 *ECMWF METEOSAT and GEOS CSR archives.*

Satellite	Time Period	Data Type	BUFR Subtype	Location
METEOSAT-5 METEOSAT-6 METEOSAT-7	15/05/1996 To ??/05/1997	Geostationary radiances, 32 × 32 pixel segments, 4 times daily	88	MARS
METEOSAT-5 METEOSAT-6 METEOSAT-7	02/05/1997 To 14/01/2002	Geostationary radiances, 32 × 32 pixel segments, Hourly	88	MARS
METEOSAT-5 METEOSAT-6 METEOSAT-7	Since 25/01/1999	Geostationary clear-sky radiances, 16 × 16 pixel segments, hourly Including clear and cloudy sky fractions	89	MARS
METEOSAT-2 METEOSAT-3	Periods for ERA	Geostationary clear-sky radiances (as above)	89	ECFS(1)
GEOS-8 GEOS-10	Since 24/10/2001	Clear sky brightness temperatures, 11 × 17 pixel segments, hourly, Including clear and cloudy sky fractions	89 and original BUFR formats Several changes	ECFS(2)
GOES-8 GOES-10	Since 09/04/2002	Clear sky brightness temperatures, 11 × 17 pixel segments, hourly, Including clear and cloudy sky fractions	89 And original BUFR formats Several changes	MARS original data on ECFS(2)

ECFS(1): ec:/ERA/era40/obs/bufr/EUM_reproc/$yyyy/$mm/CSR${yyyymmddhh}
ECFS(2): ec:/oparch/gicsbt/$yyyymm/$dd/gicsbt..

all information from the original BUFR can be retained in this BUFR and a change may be therefore useful once the incoming GOES data are encoded in the agreed common BUFR format, using descriptor 301023.

9.7.3 Thinning and screening prior to insertion into the assimilation

In order to reduce the data load of the hourly CSR data, the data are screened in a separate task before insertion into assimilation (IFS). This is done by the program GEOS_PRESCREEN (SATRAD library). It decodes the BUFR and applies basic checks on latitude, longitude, time values, and on brightness temperatures being within a physical range. Also, data points are rejected where the value for the water vapour channel brightness temperature is missing. Based on specifications given through namelist input, a geographical thinning may (or may not) be applied for each individual satellite. If switched on, the thinning is performed separately for data falling into hourly timeslots. An overview of the number of remaining valid data points per hour and satellite is printed and the remaining data are encoded into BUFR using the same format as the input file.

9.8 DEFINITIONS

9.8.1 Observation characteristics: instrument specification and retrieval type

Tables 9.11 to 9.19 describe in details how the ODB's instrument specification word is structured. Tables provided are for different observation types.

In Table 9.20 the ODB's header retrieval word codes are described.

Table 9.11 *SYNOP instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	32 – SYNOP Instrument Code Type
Not Defined	10–30	21	Reserved

Table 9.12 *AIREP instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	23 – AIREP Instrument Code Type
Flight Information	10	4	BUFR Code Table 8004 – Flight Phase
Not Defined	10–30	21	Reserved

Table 9.13 *SATOB instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	60 - GOES 62 – METEOSAT 63 – Indian SATOB 68 – Japan
I1 (Country Name)	10	4	0 – Europe 1 – Japan 2 – USA 3 – USSR 4 – India
I2I2 (Satellite Indicator Figure)	14	8	4 – METEOSAT 177 – Pretoria 0 – GEOS 3 – Japan 20 – India
Not Defined	22–30	8	Reserved

Table 9.14 *DRIBU instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	Not Defined
K1	10	4	Not Defined
K2	14	4	Not Defined
K3	18	4	Not Defined
Not Defined	22–30	8	Reserved

Table 9.15 *TEMP instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	Not Defined
Not Defined	10–30	21	Reserved

Table 9.16 *PILOT instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	Not Defined
A4	10	4	Not Defined
Not Defined	14–30	17	Reserved

Table 9.17 *SATEM instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	23	77 777 777B
I3	24	4	WMO Manual On Codes, vol II, section II-4-E-8
I4	28	4	Data processing technique. WMO Manual On Codes, vol II, section II-4-E-9
I2I2	32	7	Satellite name. WMO Manual on Codes, vol II, section II-4-E-7
I1	39	4	Country operating satellite. WMO code 1761
IS	43	7	Instrument specification code. Research Manual 5, Table 7.5
Not Defined	50	18	Reserved

9.8.2 Vertical coordinate: pressure, satellite ID and level ID codes

In the ODB the vertical coordinate is expressed by various codes, and Table 9.21 describes those codes.

Also, the ODB pressure code word is expressed in terms of codes which are defined in Table 9.22.

Each satellite used in the assimilation has is identification attached to it. The satellite identification codes used are described in Table 9.23.

Upper air observations (TEMP and PILOT) have the level at which the observation was taken defined in terms what it is and that information is stored in the ODB. Details are given in Table 9.24.

9.8.3 ODB report status: events, flags and codes

The status of each ODB report is described in terms of being active, passive, rejected or blacklisted. The ODB report status word is packed with the 4 bits given in Table 9.25.

There is one, 31 bits packed, word for each ODB report to account for various blacklist events. Details are given in Table 9.26.

Each ODB report has two words to store report events. Each report event word uses 31 bits. These events are set during observation processing to describe in more details what happened with a report.

The first ODB report event word is described in Table 9.27.

The second ODB report event word holds an additional set of events which are now dependent on observation type. Details are given in Tables 9.28 to 9.37.

Table 9.18 *TOVS instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	
A	10	2	0 – No HIRS/2 Data 1 – Clear Radiances are Derived from Clear Spots 2 – Clear Radiances are Derived from the N* method
B	12	2	0 – No HIRS/2 Data 1 – All HIRS/2 channels were used 2 – Tropospheric HIRS/2 channels were unusable due to clouds and only stratospheric channels were used
C	14	2	0 – Statistical retrieval method used 1 – Minimum information retrieval used 2 – Minimum information retrieval attempted but statistical retrieval used
V	16	3	0 – No retrieval 1 – HIRS+MSU 2 – HIRS
W	19	3	0 – No retrieval 1 – HIRS+MSU 2 – HIRS
X	22	3	0 – No retrieval 1 – HIRS(1, 2, 3, 8, 9, 16, 17)+MSU(4) 2 – HIRS(1, 2, 3, 8, 9, 16, 17) 3 – HIRS(1, 2, 3, 9, 17)+MSU(4) 4 – HIRS(1, 2, 3, 9, 17)
Y	25	3	0 - No retrieval 1 – HIRS+SSU+MSU(3, 4) 2 – HIRS+MSU(3, 4) 3 – SSU+MSU(3, 4)
Z	28	3	Not Defined
Not Defined	31	1	Reserved

Table 9.19 *SSMI instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	Not defined
Not Defined	10–31	22	Reserved

Table 9.20 *Satellite retrieval codes.*

Retrieval Codes	Description
1	Clear
2	Partly Clear
3	Cloudy

Table 9.21 *Vertical coordinate.*

Vertical Coordinate Codes	Description
1	Pressure (Pa)
2	Height (GPM)
3	Satellite Channel (numeric)
4	Scatterometer Channel (numeric)

Table 9.22 *Pressure codes.*

Pressure Codes	Description
0	Sea Level
1	Station Level
2	850 hPa Geopotential
3	700 hPa Geopotential
4	500 hPa Geopotential
5	1000 GPM Pressure
6	2000 GPM Pressure
7	3000 GPM Pressure
8	4000 GPM Pressure
9	900 hPa Geopotential
10	1000 hPa Geopotential
11	500 hPa Geopotential
12	925 hPa Geopotential

Table 9.23 *Satellite IDs.*

Satellite ID Codes	Description
208/906	NOAA10 – TOVS
235	NOAA10 – SATEM
201/907	NOAA11 – TOVS
236	NOAA11 – SATEM
202/908	NOAA12 – TOVS
237	NOAA12 – SATEM
???/909	NOAA13 – TOVS
206/910	NOAA14 – TOVS
239	NOAA14 – SATEM
205/911	NOAA15
207	NOAA16
208	NOAA17
209	NOAA18
210	NOAA19
222	NOAA20
202/241	DMSP8
203/242	DMSP9
204/243	DMSP10
205/244	DMSP11
245	DMSP12
246	DMSP13
247	DMPS14
248	DMSP15
1022	DMSP16

Table 9.24 *Level ID.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Max Wind Level
1	1	1 – Tropopause
2	1	1 – D Part
3	1	1 – C Part
4	1	1 – B Part
5	1	1 – A Part
6	1	1 – Surface Level
7	1	1 – Significant Wind Level
8	1	1 – Significant Temperature Level
9–31	24	Not Defined

Table 9.25 *Report Status.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Report Active
1	1	1 – Passive Report
2	1	1 – Rejected Report
3	1	1 – Blacklisted Report

Table 9.26 *Blacklist Events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Monthly Monitoring
1	1	1 – Constant Blacklisting
2	1	1 – Experimental Blacklisting
3	1	1 – Whitelisting
4	1	1 – Experimental Whitelisting
5	1	1 – Observation Type Blacklisting
6	1	1 – Station ID Blacklisted
7	1	1 – Code Type Blacklisted
8	1	1 – Instrument Type Blacklisted
9	1	1 – Date Blacklisted
10	1	1 – Time Blacklisted
11	1	1 – Latitude Blacklisted
12	1	1 – Longitude Blacklisted
13	1	1 – Station Altitude Blacklisted
14	1	1 – Blacklisted due to Land/Sea Mask
15	1	1 – Blacklisted due to Model Orography
16	1	1 – Blacklisted due to distance from reference point
17–30	14	Not Used

Table 9.27 *Global report events.*

Bit Position	No. of Bits	Description (Value)
0	1	1 – No Data in Report
1	1	1 – All Data Rejected
2	1	1 – Bad Reporting Practice
3	1	1 – Rejected due to RDB Flag
4	1	1 – Activated due to RDB Flag
5	1	1 - Activated by Whitelist
6	1	1 – Horizontal Position out of Range
7	1	1 – Vertical Position out of Range
8	1	1 – Time out of Range
9	1	1 – Redundant Report
10	1	1 – Over Land
11	1	1 – Over Sea
12	1	1 – Missing Station Altitude
13	1	1 – Model Surface too far from Station level
14	1	1 – Report Rejected via Namelist
15	1	1 – Failed Q/C
16–30	15	Not Used

Table 9.28 *SYNOP report events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.29 *AIREP report events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.30 *SATOB report events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.31 *DRIBU report events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.32 *TEMP report events.*

Bit Position	No. of Bits	Value – Description
0	1	1 - Old Style Z Bias Correction Applied
1	1	1 - New Style T Bias Correction Applied
2	1	1 - RH Bias Correction Applied
3–30	28	Not Used

Table 9.33 *PILOT report events.*

Bit Position	No. of Bits	Value – Description
0	1	1 - American Wind Profiler
1	1	1 - European Wind Profiler
2–30	29	Not Used

Table 9.34 *SATEM report events.*

Bit Position	No. of Bits	Value – Description
0	1	1 - Thinned Report
1–30	30	Not Used

Table 9.35 *PAOB report events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.36 *SCAT report events.*

Bit Position	No. of Bits	Value – Description
0	1	1 - Thinned Report
1	1	1 - Reported Wind Directions too Close
2	1	1 - Report not in QuikScat Sweet Spots
3	1	1 - Report Contaminated by Rain
4–30	29	Not Used

Table 9.37 *Raw radiance report events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

The ODB report RDB flag word is 30 bits packed which contains flags for five report parameters: latitude, longitude, date, time and altitude. Each parameter occupies 6 bits with further stratification which is identical for every parameter as indicated in Table 9.38.

9.8.4 Datum status: events, RDB and analysis flags

The status of each datum, like report status, is described in terms of being: active, passive, rejected or blacklisted. Table 9.39 shows that the ODB datum status is a packed word with 4 bits used to describe its status.

There are two ODB words reserved for datum events. They both use 31 bits each to store relevant information. The first event word has the same structure for all observation types, whereas the second event word is observation type dependent. Tables 9.40 to 9.50 describe the event words structures.

Furthermore, each datum in the ODB has a blacklist event word. This word uses 31 bits to describe various blacklist events as indicated in Table 9.51.

Table 9.38 *RDB report (latitude, longitude, date, time and altitude) flags.*

Parameter	No. of Bits	Bit Position	Bit Position	No. of Bits	Value – Description
			0	1	0 – No Human Monitoring Substitution 1 – Human Monitoring Substitution
Latitude	6	0+	+1	1	0 – No Q/C Substitution
Longitude	6	6+			1 – Q/C Substitution
Date	6	12+	+2	1	0 – Override Flag not Set
Time	6	18+			1 – Override Flag Set
Altitude	6	24+	+3	2	0 – Parameter Correct 1 – Parameter Probably Correct 2 – Parameter Probably Incorrect 3 – Parameter Incorrect
			+5	1	0 – Parameter Flag Set by Q/C or not Checked 1 – Parameter Flag Set by Human Monitoring

Table 9.39 *Datum status.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Report Active
1	1	1 – Passive Report
2	1	1 – Rejected Report
3	1	1 – Blacklisted Report

Table 9.40 *Global datum events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Missing Vertical Coordinate
1	1	1 – Missing Observed Value
2	1	1 – Missing Background (First Guess) Value
3	1	1 – Rejected due to RDB Flag
4	1	1 – Activated due to RDB Flag
5	1	1 – Activated by Whitelist
6	1	1 – Bad Reporting Practice
7	1	1 – Vertical Position out of Range
8	1	1 – Reference Level Position out of Range
9	1	1 – Too Big First Guess Departure
10	1	1 – Too Big Departure in Assimilation
11	1	1 – Too Big Observation Error
12	1	1 – Redundant Datum
13	1	1 – Redundant Level
14	1	1 – Report Over Land
15	1	1 – Report Over Sea
16	1	1 – Not Analysis Variable
17	1	1 – Duplicate Datum/Level
18	1	1 – Too Many Surface Data
19	1	1 – Multi Level Check
20	1	1 – Level Selection
21	1	1 – Vertical Consistency Check
22	1	1 – Vertical Coordinate Changed from Z to P
23	1	1 – Datum Rejected via Namelist
24	1	1 – Combined Flagging
25	1	1 – Datum Rejected due to Rejected Report
26	1	1 – Variational QC Performed
27	1	1 – Observation Error Increased
28–30	3	Not Used

Table 9.41 *SYNOP datum events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Bias Corrected Ps
1–30	30	Not Used

Table 9.42 *AIREP datum events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.43 *SATOB datum events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.44 *DRIBU datum events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Bias Corrected Ps
1–30	30	Not Used

Table 9.45 *TEMP datum events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Bias Corrected Value Used
1–30	30	Not Used

Table 9.46 *PILOT datum events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.47 *SATEM datum events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Not Predefined Layer
1	1	1 – Layer Formed by Thinning
2	1	1 – Layer Formed by Summing Up
3	1	1 – Channel Not Used in Analysis
4	1	1 – Overwritten by ADVAR
5–30	26	Not Used

Table 9.48 *PAOB datum events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.49 *SCAT datum events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.50 *Raw radiances datum events.*

Bit Position	No. of Bits	Value – Description
0–30	31	Not Used

Table 9.51 *Datum blacklist events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Pressure Blacklisted
1	1	1 – Variable Blacklisted
2	1	1 – Blacklisted due to Pressure Code
3	1	1 – Blacklisted due to Distance from Reference Point
4	1	1 – Blacklisted due to Type of Vertical Coordinate
5	1	1 – Blacklisted due to Observed Value
6	1	1 – Blacklisted due to First Guess departure
7–30	24	Not Used

For each datum in ODB there is an RDB flag word which holds flags for pressure (vertical coordinate) and the datum itself. This is packed word with 30 bits used – see Table 9.52. Pressure and datum RDB flags use 15 bits each. Thus pressure RDB flag starts at bit position 0, whereas the datum flag starts at bit position 15. Each 15 bits structure is further stratified in exactly the same way for both parameters:

In addition to RDB datum flags there is a word in ODB to store analysis flags. There are five types of analysis flags: final analysis, first guess, departure, variational q/c and blacklist flags. Each flag occupies 4 bits and the exact description is given in Table 9.53.

Table 9.52 *RDB pressure (vertical coordinate) and datum flags.*

Parameter	No. of Bits	Bit Position	Bit Position	No. of Bits	Value – Description
			0	1	0 – No Human Monitoring Substitution 1 – Human Monitoring Substitution
			+1	1	0 – No Q/C Substitution 1 – Q/C Substitution
Pressure	15	0+			
Datum	15	15+	+2	1	0 – Override Flag not Set 1 – Override Flag Set
			+3	2	0 – Correct 1 – Probably Correct 2 – Probably Incorrect 3 – Parameter Incorrect
			+5	1	0 – Flag Set by Q/C or not Checked 1 – Flag Set by Human Monitoring
			+6	2	0 – Previous Analysis judged it correct 1 – Previous Analysis judged it probably correct 2 – Previous Analysis judged it probably incorrect 3 – Previous Analysis judged it incorrect
			+8	1	0 – Not used by previous analysis 1 – Used by previous analysis
			+9	5	Not Used

Table 9.53 *Analysis flags.*

Flag Type	Bit Position	No. of Bits	Value – Description
Final	0	4	0 – Correct 1 – Probably correct 2 – Probably incorrect 3 – Incorrect
First Guess	4	4	0 – Correct 1 – Probably correct 2 – Probably incorrect 3 – Incorrect
Departure	8	4	0 – Correct 1 – Probably correct 2 – Probably incorrect 3 – Incorrect
Variational Q/C	12	4	0 – Correct 1 – Probably correct 2 – Probably incorrect 3 – Incorrect
Blacklist	16	4	0 – Correct 1 – Probably correct 2 – Probably incorrect 3 – Incorrect
Not Defined	20	11	Reserved

Chapter 10
Observation screening

Table of contents

10.1 INTRODUCTION

This chapter describes the observation screening in the ECMWF 3D/4D-Var data assimilation. A more general description can be found in Järvinen and Undén (1997). The purpose of the observation screening is to select a clean array of observations to be used in the data assimilation. This selection involves

quality checks, removal of duplicated observations, thinning of their resolution etc.. The current selection algorithm has been operational since September 1996 and was to a large extent designed to reproduce the functionalities of the corresponding codes in the ECMWF OI analysis (Lönnberg and Shaw, 1985, 1987; Lönnberg, 1989).

This chapter was prepared in September 1997 by Heikki Järvinen, Roger Saunders and Didier Lemeur. It was updated in February 1999 by Roger Saunders for TOVS processing, by Elias Holm and Francois Bouttier for the remainder, with further updates in October 2004 (to Cy28r1) by Erik Andersson, Drasko Vasiljevic, Tony McNally (radiance data) and Hans Hersbach (scatterometer).

10.2 THE STRUCTURE OF THE OBSERVATION SCREENING

10.2.1 The incoming observations

Before the first trajectory integration of the assimilation various observation processing steps take place. The observations for the current assimilation period are extracted from the Reports Data Base (RDB) of observations, coded in BUFR. The extracted data reside in separate BUFR files for each main observing system, e.g. conventional, ATOVS, geostationary radiances, AIRS and so on. These data have already undergone some rudimentary quality control, e.g. a check for the observation format and position, for the climatological and hydrostatic limits, as well as for the internal and temporal consistency. The so-called RDB flag has been assigned according to the outcome of such checks.

The Observation Data Base (ODB, see separate documentation) is filled from the BUFR files in a sequence of jobs called BUFR2ODB. These jobs are multi-tasked running on parallel servers. Several or all observation types can run synchronously. The resulting 'raw' ODB is processed further by the IFS, in preparation for the main analysis tasks. These ODB and data manipulation tasks are referred to as 'the make CMA replacement' (for historical reasons), and are activated by the switch LMKCMARPL. MKCMARPL entails format conversions, changes of some observed variables, such as calculation of relative humidity from dry and wet bulb temperatures and wind components from speed and direction. The assignment of observation error statistics is also done at this stage, at least for conventional data.

The resulting 'extended' ODB data base (the ECMA) contains all the observational information for the data window as required for 3D/4D-Var as well as all data that are going to be monitored. The next step is that the observations are compared to the model as it is integrated for the length of the assimilation window (Chapter 2). The observation minus model differences (the departures) are computed as described in Chapter 5 and stored in the ODB. These departures are an important input to the data selection procedures as many quality-control decisions depend on the magnitude of the departure. The collection of routines that perform data selection are jointly referred to as 'the screening'. The purpose of the observation screening is to select the best quality observations, to detect duplicates, and reduce data redundancy through thinning.

10.2.2 The screening run

The ECMWF 3D/4D-Var data assimilation system makes use of an incremental minimization scheme, as described in Chapter 2. The sequence of jobs starts with the first (high resolution) trajectory run. During this run the model counterparts for all the observations are calculated through the non-linear observation operators, and the observation minus model difference (the departures) are calculated. As soon as these background departures are available for all observations, the screening can be performed. Prior to the screening the model fields are deallocated (dealmod) as most of the information necessary in the screening is stored in the observation data base (ODB). For the observation screening, the background errors (available as grid data in the 'errgrib' file, see Chapter 14) are interpolated to the observation locations for the observed variables (INIFGER, SUFGER and GEFGER).

Technically, the final result of the observation screening is a pair of ODBs. The original 'extended' observation data base now contains observations complemented by the background departures, together with quality control information for most of the observations. This ECMA ODB remains on disc for later use in feedback creation. The compressed ODB, the CCMA, is a subset of the original observations, and

is passed for the subsequent minimization job. The CCMA contains only those observations that are to be used in the minimisation.

10.2.3 General rationale of the observation screening

The general logic in the 3D/4D-Var observation screening algorithm is to make the *independent* decisions first, i.e. the ones that do not depend on any other observations or decisions (DECIS). One example is the background quality control for one observed variable. These can be carried out in any order without affecting the result of any other independent decision. The rest of the decisions are considered as mutually *dependent* on other observations or decisions, and they are taken next, following a certain logical order. For instance, the horizontal thinning of TOVS reports is only performed for the subset of reports that passed the background quality control. Finally, the CCM data base is created for the minimization in such a way that it only contains the data that will be used.

10.2.4 3D- versus 4D-Var screening

In the original 3D-Var assimilation system the screening rules were applied once, for the complete set of observations spanning a six-hour period. In the early implementation of the 4D-Var assimilation system, the same data selection approach called '3D-screening' was applied over the 6-hour long 4D-Var time window, which resulted in essentially the same screening decisions as in 3D-Var.

In summer 1997, a new screening procedure called 4D-screening was implemented that took into account the temporal distribution of the observations. The time window is divided into time-slots of typically half-hour length (15 minutes for the first and the last time slots). The 3D-screening algorithm was then applied separately to observations within each time-slot. This allowed more data to be used by 4D-Var, for instance, all messages from an hourly reporting station can now be used, whereas only one (closest to central time) would have been allowed by the redundancy check in the 3D-screening. The 4D-screening behaviour is activated by switch LSCRE4D; it is meant to be used in conjunction with time correlation of observation errors where appropriate, as explained in Järvinen *et al.* (1999) and in Chapter 5. Also the current 3D-FGAT configuration (Chapter 3) relies on 4D-screening (LSCRE4D = .TRUE.).

10.3 THE INDEPENDENT OBSERVATION SCREENING DECISIONS

10.3.1 Preliminary check of observations

The observation screening begins with a preliminary check of the completeness of the reports (PRECH). None of the following values should be missing from a report: observed value, background departure, observation error and vertical coordinate of observation. Also a check for a missing station altitude is performed for synop, temp and pilot reports. The reporting practice for synop and temp mass observations (surface pressure and geopotential height) is checked (REPRA), as explained in Appendix A. At this stage also, the observation error for synop geopotential observations is inflated if the reported level is far from the true station level (ADDOER). The inflation is defined as a proportion of the difference between the reported level and the true station altitude by adding 2% of the height difference to the observation error.

10.3.2 Blacklisting

Next, the observations are scanned through for blacklisting (subroutine BLACK). At the set-up stage the blacklist interface is initialized (BLINIT) to the external blacklist library. The interface between the IFS and the blacklist described in further detail in Chapter 5, and in the full Blacklist documentation available elsewhere. The blacklist files (Chapter 14) consist formally of two parts. Firstly, the selection of variables for assimilation is specified in the 'data selection' part of the blacklist file. This controls which observation types, variables, vertical ranges etc. will be selected for the assimilation. Some more complicated decisions are also performed through the data selection file; for instance, an orographic rejection limit is applied in the case of the observation being too deep inside the model orography. This part of the blacklist also provides a handy tool for experimentation with the observing system, as well as with the assimilation system itself. Secondly, a 'monthly monitoring' blacklist file is provided for discarding the stations that

Table 10.1 *The predefined limits for the background quality control, given in terms of multiples of the expected variance of the normalized background departure.*

Variable	Flag 1	Flag 2	Flag 3
u, v	9.00	16.00	25.00
z, ps	12.25	25.00	36.00
dz	x	x	x
T	9.00	16.00	25.00
rh, q	9.00	16.00	25.00

Flag values are denoted by 1 for a probably correct, 2 for a probably incorrect and 3 for an incorrect observation. The variables are denoted by u and v for wind components, z for geopotential height, ps for surface pressure, dz for thickness, T for temperature, rh for relative humidity and q for specific humidity, respectively.

have recently been reporting in an excessively noisy or biased manner compared with the ECMWF background field.

10.3.3 Background quality control

The background quality control (FIRST) is performed for all the variables that are intended to be used in the assimilation. The procedure is as follows. The variance of the background departure $y - H(\chi_b)$ can be estimated as a sum of observation and background-error variances $\sigma_o^2 + \sigma_b^2$, assuming that the observation and the background errors are uncorrelated. After normalizing with σ_b, the estimate of variance for the normalized departure is given by $1 + \sigma_o^2/\sigma_b^2$. In the background quality control, the square of the normalized background departure is considered as suspect when it exceeds its expected variance more than by a predefined multiple (FGCHK, SUFGLIM). For the wind observations, the background quality control is performed simultaneously for both wind components (FGWND). In practice, there is an associated background quality-control flag with four possible values, namely 0 for a correct, 1 for a probably correct, 2 for a probably incorrect and 3 for an incorrect observation, respectively (SUSCRE0). Table 10.1 gives the predefined limits for the background quality control in terms of multiples of the expected variance of the normalized background departure. These values are set in DEFRUN and can be changed in namelist NAMJO. For SATOB winds the background error limits are modified as explained in Appendix A.

There is also a background quality control for the observed wind direction (FGWND). The predefined error limits of 60°, 90° and 120° apply for flag values 1, 2 and 3, respectively. The background quality control for the wind direction is applied only above 700 hPa for upper-air observations for wind speeds larger than 15 ms^{-1}. If the wind-direction background quality-control flag has been set to a value that is greater than or equal to 2, the background quality-control flag for the wind observations is increased by 1. For scatterometer winds, a test for high wind speeds and cold SST is applied in the IFS routine FGWND.

10.4 SCREENING OF SATELLITE RADIANCES

10.4.1 Pre-screening

Radiance observations undergo a pre-screening process before being loaded into the OBD for input to the main IFS screening. Firstly, this is used to reduce the data volume and thus the computational burden of the main screening. Secondly, this rejects observations that fail to contain crucial header information and / or the correct number of channels that could potentially cause a computational run-time failure in the main screening. Observations in BUFR are decoded and checked inside SCREEN_1C where, additionally, data measured at particular scan lines and or scan positions may be removed to reduce the data volume (by setting LINE_THIN, FOV_THIN in the calling script PRE_1CRAD). Observations which survive the

checking and thinning process are then re-encoded in BUFR and supplied to the ODB loader. A key consideration for rejecting data in the pre-screening is that removed observations will NOT be passed through the IFS screening and thus will NOT accrue feedback quality information. Currently all pre-screening tasks are scalar (i.e. not parallel). However, for IASI (by far the largest data volume) the process is effectively parallelized by splitting the input BUFR file and launching multiple scalar tasks simultaneously.

10.4.2 Cloud and rain rejection

After the pre-screening, surviving observations are passed into the main IFS screening process. Here, data contaminated by significant cloud or rain signals must be removed before being supplied to the 4D-Var minimization in the clear-sky assimilation scheme. Microwave radiances from the SSM/I(S), TMI are checked in routine RAD1CEMIS and a flag set if significant contamination is identified. The detection is based on a cloud liquid water regression algorithm . For other microwave radiances (AMSU-A/B) rain contamination is detected by scene classification based on observed window channel values inside AMSU_SFC. For infrared radiances the test for clouds is done in routine CLOUD_DETECT for AIRS/IASI and routine HIRS_CLD for HIRS. The former is based on the algorithm described in McNally and Watts 2003. The latter is described in Kelly 2007. In both cases the aim is to identify which infrared channels can be used in a particular scene and which must be rejected - neither search simply for fully clear scenes. For both the microwave and infrared data, if cloud or rain is detected and the rejection flag is set - these observations will not influence any bias evolution in the VARBC.

10.4.3 Blacklisting decisions

Like any other observations decisions are made to use or not use a particular radiance observation in the blacklist. These fall into two distinct types: The first is the usual a priori type decision which takes no account of the actual value of the observation. Examples for radiances include the exclusion of data measured by new instruments which we do not yet wish to use, data measured by bad / failed instruments, data measured at extreme scan positions, exclusion of data measured over land or high orography and the exclusion of data at certain times of year when solar intrusions may cause problems (there are others). The second type of test is particular to radiances and is a run-time decision based on the observed values (or more correctly the radiance departure f rom the background). Depending on the magnitude of the radiance departure in key window channels, individual or combinations of microwave and infrared channels may be rejected. In some respects this may be considered an additional first-guess check that takes place in the blacklists. It can equally well be considered as an additional cloud / rain detection check that takes place in the blacklist as it exclusively involves window channels. No attempt is made here to document the particular test and threshold which are applied to each channel on every instrument and the user is referred to the data selection blacklists files for details. For both types of test applied in the blacklists environment, if it is failed there are two options for what then results. The setting of a FAIL(CONSTANT) flag means that the observations will effectively be rejected and take no further part in the analysis. The setting of a FAIL(EXPERIMENTAL) flag means that the observation will enter the main analysis in such a way that it cannot force increments of e.g. temperature or humidity, but it can influence the calculation and evolution of bias correction coefficients inside VARBC. An example of when the latter is used would be for new a satellite for which we do not we wish to actively assimilate the data, but wish to establish an accurate bias correction. Another application of the F AIL(EXPERIMENTAL) facility is its use for window channels used in the quality control of other data.

10.5 SCATTEROMETER PROCESSING

10.5.1 Introduction

This section describes the flow of ERS, NSCAT, and QuikSCAT scatterometer data through the assimilation system. The retrieval of 50 km QuikSCAT winds and some tasks for ERS are performed in the module PRESCAT, which, being part of the OBS family, are run before the screening. But it is most natural to describe the whole processing step here. This section provides a working knowledge of the software, and guidance on possible modifications and updates. It is not intended to explain the scientific

background of microwave remote sensing or scatterometry and assumes some knowledge of these topics (see Stoffelen, 1999).

Besides the introduction, this section is broken into five subsections. The next subsection gives background information on scatterometer processing at ECMWF. It includes a brief description of the various instruments, their period of functioning and operational assimilation at ECMWF. Besides it provides some history of the software, lists persons who contributed to the changes and outlines the structure and function of the whole library. Subsection 10.5.3 gives concise information on the location of scatterometer data in the ECMWF archives. Subsections 10.5.4 to 10.5.6 describe the line of processing for scatterometers (used or) currently in use at ECMWF, i.e. (ERS-1 and) ERS-2, (NSCAT) and QuikSCAT.

10.5.2 Background and general scatterometer geometry

ESA's ERS-1 scatterometer was launched in July 1991 and stopped operating in June 2000. Its successor, ERS-2, was launched in 1995 and is still working. However, due to the loss of a proper functioning of its gyroscopes in January 2001 the yaw attitude control of the spacecraft became deteriorated. This especially affected the quality of scatterometer data, and as a result, its worldwide distribution was suspended. However, these problems were overcome by the combination of alternative on-board attitude control and the introduction of a completely revised on-ground processor. As a result, ESA resumed data distribution on 21 August 2003. Due to the loss of the on-board low-bit-rate recorders a few months earlier, global coverage was lost and data coverage is since then confined to the North-Atlantic and part of the Mediterranean. Data from ERS-2 were introduced into operations at ECMWF in January 1996, have been used until the suspension in January 2001, and were re-introduced on 9 March 2004. The NSCAT instrument was part of the payload of the ADEOS-I satellite, which was launched in August 1996. NSCAT data have been available from mid September 1996 until the spacecraft failure on June 30 1997. NSCAT data have never been part of the operational 4-D Var system. The QuikSCAT satellite, carrying the SeaWinds scatterometer instrument, was launched in June 1999. Its data was introduced in the operational 4D-Var system on 22 January 2002 and has been assimilated since that date. A second SeaWinds instrument, being on-board the Midori-II satellite, was launched in December 2002. Unfortunately, the Midori-II mission was lost in October 2003, which occurred just a few weeks before the planned public release of the SeaWinds data. Consequently data from this instrument was never assimilated at ECMWF.

The scatterometers aboard the ERS-1 and ERS-2 satellites are of identical design. Operating at C-band frequency, three antennas at the right-hand side from the flight direction (45°, 90° and 135°) illuminate the ocean surface. Inside the 450 km wide swath, 19 nodes define a 25 km product. For each such node, average backscatter values for the three beams are determined from a 50 km by 50 km footprint, i.e., the 25 km product is over-sampled. More detailed information on the ERS scatterometers can be found in Attema (1986), its impact on the ECMWF assimilation system in Isaksen (1997) and Isaksen and Stoffelen (2000).

The NSCAT scatterometer that was part of the payload of the ADEOS-I satellite, consisted of three antennas on each side of the spacecraft. It operated at Ku-band frequency, and provided coverage in two simultaneous 600 km wide swaths, separated by a 350 km wide nadir gap. The resolution of the backscatter and wind product was 50 km. More detailed information on the NSCAT scatterometer can be found in the appendix of Chelton *et al.* (2000), its impact on the ECMWF assimilation system in Leidner *et al.* (2003).

The Seawinds scatterometer aboard the QuikSCAT and Midori-II satellites, finally, consists of two rotating beams operating at Ku-band frequency with incidence angles w.r.t. the normal of the ocean surface of 46° respectively 56°. Each beam provides a fore and an aft measurement. They cover a swath of 1,800 km in diameter, although only the inner 1,400 km is illuminated by both beams. The product is determined on a 25 km resolution, defining 76 across nodes. The first eight and last eight nodes contain two backscatter measurements (outer beam only), the other nodes can have up to 4 backscatter measurements. More detailed information on the QuikSCAT scatterometer can be found in Leidner *et al.* (2000), its impact on the ECMWF assimilation system in Hersbach *et al.* (2004).

Ad Stoffelen, David Anderson and Ross Hoffman were the first to work on the problem at ECMWF. Stoffelen and Anderson worked on quality control (QC) and wind retrieval issues in the OI system

Table 10.2 *The MARS archive definitions.*

	BUFR/MARS obstype	CMA code type	Satellite ID
ERS-1	122	122	1
ERS-2	122	122	2
NSCAT	136	210	280
QuikSCAT	137*	300	281
QuikSCAT	138*	301	281

*For QuikSCAT the BUFR format changed in February 2000 which is reflected in the change of MARS obstype from 137 to 138 and the change in the BUFR sequence descriptor.

of these days. Hoffman looked at assimilation of backscatter information directly in 3D-Var. Once in operations, several others (Herve Roquet, Catherine Gaffard, Didier LeMeur and Lars Isaksen) took turns monitoring and improving the use of the data. Later Mark Leidner worked on the use of data from NASA scatterometers (NSCAT and QuikSCAT) and the operational implementation of the assimilation of QuikSCAT data was completed by Hans Hersbach and Lars Isaksen. The re-introduction of ERS-2 scat data in March 2004, and the formulation of an improved geophysical model function CMOD5, now being used for the inversion of ERS backscatter triplets to winds, was performed by Hans Hersbach.

Source code for scatterometer processing resides in ClearCase under the project name SCAT. The library contains the following directories.

(i) **etimesort**/ source code for pre-processing ERS data.
(ii) **module**/ shared modules.
(iii) **qbukey**/ source code for adding RDB info to QuikSCAT 50 km BUFR.
(iv) **qfilter**/ source code for pre-processing QuikSCAT 25 km BUFR.
(v) **qretrieve**/ source code for SeaWinds wind retrieval.
(vi) **dcone_qc**/ source code to perform a global quality control check on QuikSCAT data with the goal to detect instrument anomalies.
(vii) **qsca_split**/ source code that allows for the distribution of QuikSCAT pre-processing over several processors (not active).
(viii) **monitoring**/ empty directory in which scatterometer monitor software may reside in future.
(ix) **test**/ empty directory for future test code.

NSCAT-specific codes have not been put in ClearCase, because the satellite stopped operating in June 1997 (i.e. its data will never be used in operations).

10.5.3 Location of scatterometer data in the ECMWF archives

The analysis input data for both ERS and QuikSCAT are available from MARS in BUFR format. This is not the case for NSCAT data. Data for the entire NSCAT 9-month mission are stored on ecfs in HDF format at the following.

(i) ecfs:/oparch/nscat/50km/L17 - Level 1.7 files (backscatter data).
(ii) ecfs:/oparch/nscat/50km/L20 - Level 2.0 files (wind data), instead.

No SeaWinds data from Midori-II are present in the ECMWF archives.

The MARS archive definitions for the different wind scatterometer observations are given in Table 10.2.

10.5.4 ESA scatterometer (ERS-1 and ERS-2) processing

Fig. 10.1 shows a simple flow chart for ERS processing at ECMWF. Below the processing chain is described in general and the functionality of each executable in the scat library in particular.

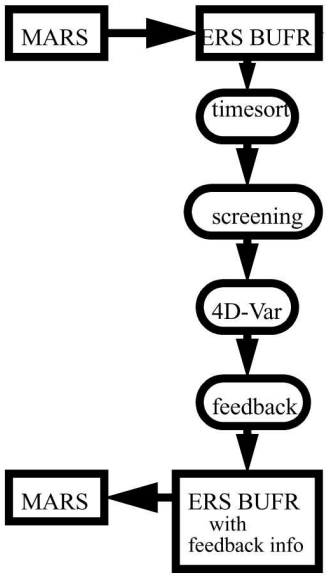

Figure 10.1 *ERS processing.*

Data for a given time window (12 hours in operations) are retrieved from MARS into patches of 6 hours, and time-wise centered around 00H, 06H, 12h and 18H. These data are then input to program TIMESORT, which sorts the observations by time and removes duplicate records. Duplicate records occur in the input data because more than one ground station may receive the same ERS data. Duplicated data are almost identical and their time stamps will differ by a small margin (<4 seconds). The software for removing duplicates, or occasionally even triplets, resides in **etimesort**. The original algorithm was redesigned in Cy28r1.

ERS winds are retrieved as part of the IFS observation pre-processor (see Chapter 9). Here, the three backscatter measurements are decoded from BUFR, and used to retrieve winds for 50 km diameter foot prints (the ERS winds are available at an over-sampled 25 km resolution). It is not possible to determine a unique wind vector solution, at least two ambiguous solutions will be found. Only the most likely wind and the first one in the opposite direction are kept and written to the observation data base (ODB) file. For this ERS winds are retrieved in-house, rather than using the unique wind distributed by ESA, to be able to supply two winds to the variational data assimilation system at ECMWF, and to be able to apply bias corrections to backscatter values before wind retrieval, and finally to be able to apply quality controls to the retrieved winds via the retrieval residual.

The in 25 km-resolution available ERS data is subject to horizontal thinning, and is controlled by LSCATTHI and NTHINSCA in namelist NAMMKCMA. In the across-swath direction the inner two nodes are skipped (smallest incidence angles) because the scatterometer operates best at larger incidence angles. Then, every NTHINSCA node is used (default NTHINSCA = 4 results in use of node 3, 7, 11, 15, and 19). Along swath, the generic thinning procedure as described in Subsection 10.6.6 (minimum distance (NTHINSCA/4)×50 km and maximum distance of (NTHINSCA/4)×100 km) is applied. As a result, the data are used at a resolution of 100 km instead of the original 25 km sampling distance. It should be noted that the thinning process is actually set up in the observation pre-processing (IFS), but that only a flag is assigned at that stage, which is then applied in the screening. In that way all the data are completely pre-processed before assimilation, and the subsequent information kept in the feedback files allows to perform their monitoring at the full resolution.

In the IFS, both retrieved winds are used in the assimilation system. In 3D-Var, an observation cost function with two minima (see **pp_obs**/HJO) is used. This function is not quadratic, but depends on the fourth power and root of the product of the departures of u and v (Stoffelen and Anderson, 1997) so that

$$J(\mathbf{u}) = \frac{J_1 J_2}{(J_1^4 + J_2^4)^{0.25}}, \quad J_{1,2}(\mathbf{u}) = \frac{\|\mathbf{u} - \mathbf{u}_{1,2}\|^2}{\sigma_0^2}. \tag{10.1}$$

In the vicinity of each solution, this cost function looks quadratic. The observation error σ_0 is 2 m/s (defined in the **obs_preproc**/QSCATIN routine). Since the introduction (CY25R3) of the conjugate gradient minimisation technique, such a non-quadratic cost function is not longer allowed in 4D-Var. Instead, cost function (10.1) is replaced by

$$J(u) = J_{1.2}(u) \tag{10.2}$$

where $J_{1.2}$ corresponds to the wind solution that is closest to the high-resolution trajectory. As a result, both retrieved wind solutions are still used. However, de-aliasing is now performed in the outer loop, enabling for a purely quadratic cost function in the inner loop. The choice between cost function (10.1) and (10.2) is determined by the logical LQSCATT (.FALSE. for cost (10.1), .TRUE. for expression (10.2)).

Quality control decisions made by the IFS screening run are:

(i) **High wind speed check:** Data rejected if observed or first guess wind speeds are above 35 ms^{-1} (RSCAWLI). Performed by **obs_preproc**/fgwnd..
(ii) **Sea ice check:** Data rejected if sea-ice fraction is higher than 0.01 (RSCATLI). Performed by **obs_preproc**/fgwnd..
(iii) **Global quality control:** If the average distance-to-the-cone residual for the backscatter measurements during an observation time slot (30 min.) for any of the active nodes is above the QC threshold, all ERS data for that time slot is blacklisted. This is done by the routine **obs_preproc**/scaqc..

There is no background wind check performed on ERS scatterometer data, but data may be de-weighted or effectively removed from the analysis during the minimization in 4D-Var by variational quality control (see Section 2.6 or Andersson and Järvinen (1999)).

Quality control decisions and departures from background and analyses are appended to each subset in the feedback BUFR message.

ERS feedback messages have a PRESCAT section sandwiched between the original ERS and the feedback data. The PRESCAT section contains outputs about the quality of the winds from the retrieval.

Here are the ERS-specific key words and bits to examine in the feedback message.

(i) **Winds retrieved at ESA:** BUFR descriptor 11012 for speed and 11011 for direction winds available in observation part of BUFR file.
(ii) **Winds retrieved at ECMWF:** BUFR descriptor 11192 for u and 11193 for v winds retrieved in **obs_preproc**/ERSIN..
(iii) **Background departures x 2 ambiguities:** BUFR descriptor 224255 for u ('U – COMPONENT AT 10 M') and for v ('V – COMPONENT AT 10 M'). BUFR descriptor 8024 = 33, BUFR descriptor 33210 = 1, BUFR descriptor 33211 = 1001.
(iv) **Analysis departures x 2 ambiguities:** BUFR descriptor 224255 for u ('U – COMPONENT AT 10 M') and for v ('V – COMPONENT AT 10 M'). BUFR descriptor 8024 = 33, BUFR descriptor 33210 = 9, BUFR descriptor 33211 = 999.
(v) **Report rejected by across-swath thinning** if bit 1 of BUFR descriptor 33229 (Report Event Word 2) is set. QC decision made in **obs_preproc**/SCAQC..
(vi) **Report rejected by along-swath thinning** if bit 10 of BUFR descriptor 33220 (Report Event Word 1) is set. QC decision made in **obs_preproc**/THINNER..
(vii) **Report rejected if Sea Ice faction too high:** if bit 12 of BUFR descriptor 33220 (Report Event Word 1) is set. QC decision made by program IFS in subroutine **obs_preproc**/fgwnd..
(viii) **Report rejected by high wind speed check** if bit 16 of BUFR descriptor 33220 (Report Event Word 1) is set. QC decision made in **obs_preproc**/fgwnd..
(ix) **Report rejected if global QC fails:** if bit 16 of BUFR descriptor 33220 (Report Event Word 1) is set. QC decision made in **obs_preproc**/ERSQC. Note that currently this bit can be set for two different reasons, which is to be resolved in next model cycle.
(x) **Datum 4D-Var quality control status:** BUFR descriptor 33233 (Report Status Word 1) = 1/2/4/8 1 – active, 2 – passive, 4 – rejected, 8 – blacklisted Datum. 4D-Var variational quality control rejection BUFR descriptor 33236 (Datum event Word 1) bit 27 = 1.

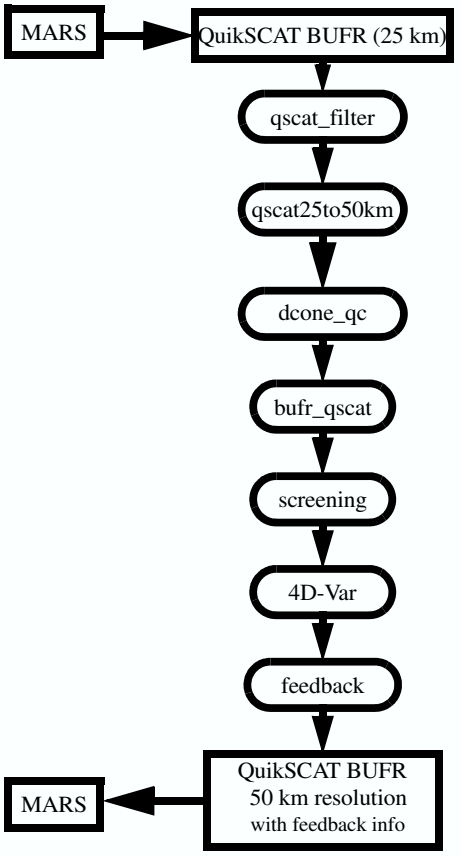

Figure 10.2 *QuikSCAT processing.*

10.5.5 NASA scatterometer (NSCAT) processing

NSCAT data has been used experimentally for impact experiments in 4D-Var as well as a surrogate for QuikSCAT data (Ku-band scatterometer as well). The processing is not automatic in IFS, as is the case for ERS and QuikSCAT. The NSCAT data quality is more consistent compared to ERS and QuikSCAT, because the archived NSCAT data are a re-processed science product, not an "as-is" real-time product.

The format and content of HDF NSCAT files are thoroughly documented in QuikSCAT Science Data Product User's Manual (available from ECMWF or JPL). Level 1.7 and 2.0 files are present for each orbit in the mission. Each backscatter file has a corresponding wind file.

Assimilation experiments with NSCAT data are only possible after offline processing of the data. Please contact the research department for further information.

10.5.6 NASA scatterometer (QuikSCAT) processing

Fig. 10.2 shows a simple flow chart for QuikSCAT processing at ECMWF. Below the processing chain is described in general. The processes QSCAT_FILTER, QSCAT25TO50KM, DCONE_QC and QSCAT_BUFR are part of the SCAT library. They are performed in the module PRESCAT, which is part of the OBS family. The processes SCREENING, and 4D-VAR, being part of the IFS library, are performed in the AN family. The FEEDBACK process, being part of the ODB library, is performed in the LAG family.

QuikSCAT data are received in the SeaWinds Real-Time BUFR Data Product (SWS_Met BUFR) format. This level 2B product contains both backscatter and wind information. The provided wind product is not used, however. Instead, wind information is retrieved directly by inversion of the backscatter product (see IFS Documentation Part I "Observation processing" Chapter 10). A detailed description of the SWS_Met

BUFR data content can be found in Leidner *et al.* (2000). The SWS_Met BUFR product is archived in MARS as indicated in Subsection 10.5.3.

Data for a given time window (12 hours in operations) are retrieved from MARS into six-hourly patches, time centered around 00H, 06H, 12h and 18H. These data are input to the program **qfilter**/QSCAT_FILTER, which sorts the observations by time, and removes duplicate/incomplete records. Duplicate and incomplete records are part of the QuikSCAT real-time data stream because of Seawinds' conically-scanning geometry. See Leidner *et al.* (1999) for a discussion of duplicate and incomplete records introduced by the scanning geometry.

QuikSCAT winds are inverted from the backscatter information with the program **qretrieve**/QSCAT25TO50KM. In contrast to ERS scatterometer data (see above), where the inversion from backscatter values to winds are defined in the IFS library, the wind inversion is defined in the SCAT library. Input to **qretrieve**/QSCAT25TO50KM are the sorted and undoubled 25 km QuikSCAT BUFR messages. They are decoded, consecutive rows are paired together, backscatter measurements are grouped into 50 km boxes, and winds are retrieved at this resolution. This means that from the 76 across track 25 km vector cells, 38 vector cells at 50 km resolution are created. Motivation for this transformation is that a 50 km resolution is more representative to the scales that are resolved by the increments in 4D-Var.

The 50 km wind product is based on all backscatter measurements available within that cell, i.e., four 25 km sub-cells. For this a Maximum Likelihood Estimator (MLE) is optimized expressing the misfit between observed and modelled backscatter values (based on the QSCAT-1 geophysical model function). Some details are described in the IFS Documentation (Part I "Observation processing" Section 10.2). The sum contains up to 16 backscatter measurements. In practice this number may be smaller because of the following.

(i) Only outer-beam measurements are available (up to 8 values, for nodes 1–4 and 35–38).
(ii) Missing backscatter values.
(iii) Missing Kp values (see Leidner *et al.*, 2000).
(iv) The land, sea, and/or ice flag available in the product is set.
(v) Rain contamination.

More details on the extraction of rain-contamination information can be found in Subsection 9.6.6 Instead of the rejection on the basis of rain contamination, the code has been prepared (**qretrieve**/DIST2CONE) to reject backscatter values on the basis of a distance-to-cone method developed by Portabella and Stoffelen (2000). At the moment, such a replacement requires a (small) modification of the SCAT library in ClearCase.

The inversion is performed in the routine **qretrieve**/INVERT50. On output it returns up to four ambiguous wind solutions. These winds are bias-corrected (for details see Subsection 9.6.8) in such a way that on average, QuikSCAT wind speeds are in line with their ECMWF first-guess collocated counterparts.

On output of **qretrieve**/QSCAT25TO50KM, a 50 km BUFR file is created. Among other quantities, it contains speed and direction of all wind solutions (ordered from most likely to least likely), the wind speed biases that were applied, all 16 underlying backscatter measurements, values of the resulting MLE's, rain flag and distance-to-cone information of the underlying 25 km wave vector cells.

The in this way created 50 km BUFR six-hourly files are input to the program **dcone_qc**/DCONE_QC. Goal is the detection of possible instrument anomalies. Typically when such anomalies occur, the set of backscatter values (up to 16 for the 50 km product) does not agree well with any set of backscatter values based on the geophysical model function. As a result, the MLE for the most-likely wind solution obtained in **qretrieve**/INVERT50 will be too large. Occasionally large values can be caused by statistical fluctuations or by local geophysical parameters not taken into account by the model function. Consistently large values, however, indicate an instrument anomaly. In **dcone_qc**/DCONE_QC the across-node wise average MLE for each six-hour data file is compared to what should be expected on average. This average value is a factorization over expectation values as function of across node number, the wind speed of the most-likely solution and the number of backscatter values used in **qretrieve**/INVERT50. These

normalizing values were determined for the period between 18 UTC on 31 August 2001 and 12 UTC on 31 September 2001 and are defined in **dcone_qc**/INIT. If any of the normalized averages for nodes for node 5 to 34 (i.e., where both outer and inner beam measurements are available) is larger than 1.45, the entire six-hour BUFR file is suspect. In such case bit 7 of the WVC Quality Flag (21109) is set for all observations. Otherwise, the BUFR file is not modified. This global quality control was introduced in the ECMWF assimilation system on 14 April 2002. During the three-months period before its introduction, in which QuikSCAT data was assimilated without a global QC check, it would only have been active for the six-hourly files of 12 UTC on 12 March 2002 and 12 UTC on 19 March 2002. For diagnostic reasons, node-wise normalized averages and all observations for which the normalized distance to the cone exceeds 25, are printed to the obs/prescat.1 logfile. The reason why this quality control acts on six-hourly files, rather than on 30-minute time slots, as it is performed for ERS scatterometer data, is that the natural fluctuations of the MLE are larger for QuikSCAT data than they are for ERS data. As a result, 30-minute averages were found to be too short.

The **qbukey**/BUFR_QSCAT program adds RDB information to the 50 km BUFR file.

In IFS, ambiguous winds are reordered (most likely solution, solution that is most anti-parallel to the most likely, and then the remaining two in order of likeliness). Only the first two winds are input to 4D-Var using the same cost function as for ERS (i.e., the double well or purely quadratic cost function (10.1) respectively (10.2)), including $\sigma_0 = 2$ m/s, see also Stoffelen and Anderson, 1997. Only difference with ERS is, that the weight of the cost is reduced by a factor of four. In this way, QuikSCAT data is effectively assimilated on a resolution of 100 km. Calculation of the cost occurs in the IFS routine **pp_obs**/hjo.

Quality control decisions made by the IFS screening run are as follows.

(i) **Sea ice check:** Data rejected if sea-ice fraction is higher than 0.01 (RSCATLI). Performed by **obs_preproc**/fgwnd.
(ii) **Sweet zone check:** Data rejected if it is in the region of the swath for which no inner beam measurement is available (node 1–4 and node 35–38). This is done by the routine **obs_preproc**/scaqc.
(iii) **Insufficient number of ambiguities:** Data reject if less than two ambiguous wind solutions are available (**obs_preproc**/scaqc).
(iv) **Lack of azimuth diversity:** Data reject when the angle between the first and second reordered solution is less than 135 degrees. This mainly occurs for observations in the nadir part of the swath. This is done by the routine **obs_preproc**/scaqc.
(v) **Rain contamination:** Data reject if more than one 25 km sub-cell (from which the 50 km cell is constructed) is rain contaminated (see Section 9.3 for details). This is done by the routine **obs_preproc**/scaqc.
(vi) **Global QC:** Data reject if the global-performance flag (see **dcone_qc**/DCONE_QC program above) is set. This is done by the routine **obs_preproc**/scaqc.

In addition the **obs_preproc**/scaqc. code has been prepared to reject data in case for more than one 25 km sub-cell the normalized distance-to-the-cone is too large (see the program **qretrieve**/QSCAT25TO50KM, and Portabella and Stoffelen (2000) for its description). It is not active at the moment.

There is no background wind check performed on QuikSCAT scatterometer data, but data may be de-weighted or effectively removed from the analysis during the minimization in 4D-Var by variational quality control (Section 2.6 or Andersson and Järvinen, 1999).

After the processing in 4D-Var, a 50 km feedback file is created. Starting from the 50 km BUFR file created in the **qbukey**/BUFR_QSCAT program, it adds information obtained during the 4D-Var process. Here the QuikSCAT-specific key words and bits to examine in the feedback message.

(i) **Background departures x 4 ambiguities:** Like for ERS described above. Only the first two departures have been used in the 4D-Var cost function.
(ii) **Analysis departures x 4 ambiguities:** Like for ERS described above. Only the first two departures have been used in the 4D-Var cost function.

(iii) **Report rejected if sea ice fraction is too high:** Like for ERS described above.
(iv) **Report rejected if number of winds is < 2:** if bit 3 of BUFR descriptor 33220 (Report Event Word 1) is set. QC decision made in **obs_preproc**/scaqc..
(v) **Report rejected if wind directions are too close:** if bit 2 of BUFR descriptor 33229 (Report Event Word 2) is set. QC decision made in **obs_preproc**/scaqc..
(vi) **Report rejected if data not in the sweet zone:** if bit 3 of BUFR descriptor 33229 (Report Event Word 2) is set. QC decision made in **obs_preproc**/scaqc..
(vii) **Report rejected if data is rain contaminated:** if bit 4 of BUFR descriptor 33229 (Report Event Word 2) is set. QC decision made in **obs_preproc**/scaqc..
(viii) **25km normalized distance-to-the cone is too large (rejection is not active):** if bit 16 of BUFR descriptor 33229 (Report Event Word 2) is set. Decision made in **obs_preproc**/scaqc..
(ix) **Report rejected if global QC fails:** if bit 16 of BUFR descriptor 33229 (Report Event Word 1) is set. QC decision made in **obs_preproc**/scaqc..
(x) **Datum 4D-Var quality control:** Like for ERS described above.

10.6 THE DEPENDENT OBSERVATION SCREENING DECISIONS

10.6.1 Update of the observations

Just before performing the dependent screening decisions, the flag information gathered so far is converted into a status of the reports, namely: **active, passive, rejected** or **blacklisted**, and also into a status of the data in the reports (FLGTST). The reports with a RDB report flag value 2 (probably incorrect) or higher for latitude, longitude, date and time are rejected. For the observed data there are RDB datum flags for the variable and for the pressure, i.e. the pressure level of the observation. The rejection limits for these are as follows: all data are rejected for the maximum RDB datum flag value 3 (incorrect), non-standard-level data are rejected for the maximum RDB datum flag value 2, and for the pressure RDB datum flag the rejection limit is 1 (probably correct). The background quality control rejection limits are flag value 3 for all the data, and flag value 2 for the non-standard-level data.

10.6.2 Global time–location arrays

Some of the dependent decisions require a global view to the data which is not available as the memory is distributed. Therefore *ad hoc* global time–location arrays are formed and broadcast in order to provide this view (GLOBA, DISTR).

10.6.3 Vertical consistency of multilevel reports

The first dependent decisions are the vertical-consistency check of multilevel reports (VERCO), and the removal of duplicated levels from the reports. The vertical-consistency check of multilevel reports is applied in such a way that if four consecutive layers are found to be of suspicious quality, even having a flag value one, then these layers are rejected, and also all the layers above these four are rejected in the case of geopotential observations. These decisions clearly require the quality-control information, and they are therefore 'dependent' on the preceding decisions.

10.6.4 Removal of duplicated reports

The duplicated reports will be removed next. That is performed (MISCE, DUPLI, REDSL) by searching pairs of collocated reports of the same observation types, and then checking the content of these reports. It may, for instance, happen that an airep report is formally duplicated by having a slightly different station identifier but with the observed variables inside these reports being exactly the same, or partially duplicated. The pair-wise checking of duplicates results in a rejection of some or all of the content of one of the reports.

10.6.5 Redundancy check

The redundancy check of the reports, together with the level selection of multi-level reports, is performed next for the active reports that are collocated and that originate from the same station (REDUN).

In 3D-screening, this check applies to the whole observation time window. In 4D-screening (LSCRE4D = .TRUE.), this check applies separately in each timeslot.

For **land synop and paob** reports, the report closest to the analysis time with most active data is retained, whereas the other reports from that station are considered as redundant and are therefore rejected from the assimilation (REDRP, REDMO). For **ship synop and dribu** observations the redundancy check is done in a slightly modified fashion (REDGL). These observations are considered as potentially redundant if the moving platforms are within a circle with a radius of 1° latitude. Also in this case only the report closest to the analysis time with most active data is retained. All the data from the multilevel **temp and pilot** reports from same station are considered at the same time in the redundancy check (REDOR, SELEC). The principle is to retain the best quality data in the vicinity of standard levels and closest to the analysis time. One such datum will, however, only be retained in one of the reports. A wind observation, for instance, from a sounding station may therefore be retained either in a temp or in a pilot report, depending on which one happens to be of a better quality. A **synop mass** observation, if made at the same time and at the same station as the temp report, is redundant if there are any temp geopotential height observations that are no more than 50 hPa above the synop mass observation (REDSM).

10.6.6 Thinning

Finally, a horizontal thinning is performed for the AIREP, TOVS, GEOS, SSM/I, SATOB and ERS SCAT reports. The horizontal thinning of reports means that a predefined minimum horizontal distance between the nearby reports from the same platform is enforced. For AIREP reports the free distance between reports is currently enforced to about 60 km (Cardinali *et al.*, 2003). The thinning of the AIREP data is performed with respect to one aircraft at a time (MOVPL, THIAIR). Reports from different aircraft may however be very close to each other. In this removal of redundant reports the best quality data is retained as the preceding quality controls are taken into account. In vertical, the thinning is performed for layers around model levels, thus allowing more reports for ascending and descending flight paths.

Thinning of TOVS, GRAD, SSM/I, SATOB and ERS SCAT reports are each done in two stages controlled by THINN. For TOVS (THINNER), a minimum distance of about 70 km is enforced and, thereafter, a repeat scan is performed to achieve the final separation of roughly 250 km or 120 km between reports from one platform. This is controlled through settings in DEFRUN, that can also be modified through namelist (NAMSCC). The thinning algorithm is the same as used for AIREPs except that for TOVS a different preference order is applied: a sea sounding is preferred over a land one, a clear sounding is preferred over a cloudy one and, finally, the closest observation time to the analysis time is preferred. For geostationary water vapour radiances, a similar thinning in two stages is applied with currently about 70 km minimum distance and about 125 km final separation (THINNER). During the thinning, preference is given to data having the largest fraction of clear sky in the clear-sky radiance average, high infrared brightness temperature (for GOES data) and, finally, a small standard deviation of brightness temperatures within the CSR mean. A similar thinning technique is applied to SSM/I data and SATOB high-density data (THINNER). Note that prior to assimilation a coarser pre-thinning may take place already during observation pre-processing in order to reduce otherwise excessive data volumes.

The screening of SATOB data has been extended for atmospheric motion wind observations, including individual quality estimate. The quality information from the quality control performed by the producer at extraction time is appended to each wind observation. This Quality Indicator (QI) is introduced as an additional criterion in the thinning step; priority is given to the observation with the highest QI value.

Apart from this thinning, the other observation dependent decisions involved by the screening of the scatterometer data come essentially from the application of a sea-ice contamination test from the model sea-surface-temperature analysis, using a minimum threshold of 273 K, and a high-wind rejection test with an upper-wind speed limit set to 35 ms^{-1} for the higher of the scatterometer and background winds (FGWND).

Table 10.3 *A summary of the current use of observations in the 3D/4D-Var data assimilation at ECMWF. These usage rules are subject to change. For the most current information we refer to the data-selection blacklist files. These files are archived, so the history is preserved.*

Observation type	Variables used	Remarks
SYNOP, SHIP	u, v, ps (or z), rh	u and v used only over sea, in the tropics also over low terrain (<150 m). Orographic rejection limit 6 hPa for rh, 100 hPa for z and 800 m for ps.
AIREP	u, v, T	Thinned to approximately 60 km along flight tracks. Ascending and descending aircraft are thinned in the vertical to the model resolution.
SATOB	u, v	Selected areas and levels. Thinning of high-density winds. Strict background check to avoid slow wind-speed bias in the data.
DRIBU	u, v, ps	Orographic rejection limit 800 m for ps.
TEMP	u, v, T, q	Used on all reported levels. q only below 300 hPa. 10 m u and v used over land only in tropics over low terrain (<150 m). Orographic rejection limit 10 hPa for u, v and T, and -4 hPa for q.
PILOT	u, v	Used on all reported levels. 10 m u and v used over land only in tropics over low terrain (<150 m). Orographic rejection limit 10 hPa for u and v.
Profilers	u, v	Thinned to 5 hPa separation in the vertical. Data from American, European and Japanese networks are used.
Radiance data	Tb	Information to be supplied by the Satellite Section.
PAOB	ps	Used south of 19^0S. Orographic rejection limit 800 m for ps.
Scatterometer	u, v	Not used in full resolution. Used if SST is warmer than 273 K or if both observed and background wind less than 35 ms^{-1}.

The variables are as in Table 10.1, with the addition that Tb stands for brightness temperature. The observation types are shortened by *SYNOP* for synoptic surface observations, *AIREP* for aircraft reports, *SATOB* for satellite cloud track winds, *DRIBU* for drifting buoy reports, *TEMP* for radiosonde soundings, *PILOT* for wind soundings and *PAOB* for pseudo observations of surface pressure obtained from satellite images.

10.6.7 A summary of the current use of observations

A summary of the current status of use of observations in the 3D-Var data assimilation is given in Table 10.3. For most current information we refer to the data-selection blacklist files themselves. The history of such files as used in ECMWF operations is also available.

10.6.8 Compression of the ODB

After the observation screening roughly a fraction of 1/10 of all the observed data are active and so the compressed observation ODB (the CCMA) for the minimization run only contains those data. The large compression rate is mainly driven by the number of TOVS data, since after the screening there are only 10–20% of the TOVS reports left, whereas for the conventional observations the figure is around 40%. As a part of the compression, the observations are re-sorted amongst the processors for the minimization job in order to achieve a more optimal load balancing of the parallel computer.

10.7 A MASSIVELY-PARALLEL COMPUTING ENVIRONMENT

The migration of operational codes at the ECMWF to support a massively-parallel computing environment has set a requirement for reproducibility. The observation screening needs to result in exactly

the same selection of observations when different numbers of processors are used for the computations. As mentioned earlier, in the observation screening there are the two basic types of decision to be made. Independent decisions, on one hand, are those where no information concerning any other observation or decision is needed. In a parallel-computing environment these decisions can be happily made by different processors fully in parallel. For dependent decisions, on the other hand, a global view of the observations is needed which implies that some communication between the processors is required. The observation array is, however, far too large to be copied for each individual processor. Therefore, the implementation of observation screening at the ECMWF is such that only the minimum necessary information concerning the reports is communicated globally.

The global view of the observations is provided in the form of a global 'time–location' array for selected observation types. That array contains compact information concerning the reports that are still active at this stage. For instance, the observation time, location and station identifier as well as the owner processor of that report are included. The time–location array is composed at each processor locally and then collected for merging and redistribution to each processor. After the redistribution, the array is sorted locally within the processors according to the unique sequence number. Thus, every processor has exactly the same information to start with, and the dependent decisions can be performed in a reproducible manner independently of the computer configuration.

The time–location array is just large enough for all the dependent decisions, except for the redundancy checking of the multilevel temp and pilot reports. This is a special case, in the sense that the information concerning each and every observed variable from each level is needed. Hence, the whole multilevel report has to be communicated. The alternative to this would be to force the observation clusters of the multilevel reports always into one processor without splitting them. In that case the codes responsible for the creation of the observation arrays for assimilation would need to ensure the geographical integrity of the observation arrays distributed amongst the processors. This is, however, not possible in all the cases, and the observation screening has to be able to cope with this. Currently, it is coded in such a way that only a limited number of multilevel temp and pilot reports, based on the time–location array, are communicated between the appropriate processors as copies of these common stations.

APPENDIX A

A.1 Bad reporting practice of synop and temp reports

The way the synoptic surface stations report mass observations (pressure or geopotential height) is considered as bad if:

- station altitude is above 800 m and station reports mean sea level pressure
- station altitude is above 800 m and station reports 1000 hPa level
- station altitude is above 1700 m and station reports 900 hPa level
- station altitude is below 300 m and station reports 900 hPa level
- station altitude is above 2300 m and station reports 850 hPa level
- station altitude is below 800 m and station reports 850 hPa level
- station altitude is above 3700 m and station reports 700 hPa level
- station altitude is below 2300 m and station reports 700 hPa level
- station altitude is below 3700 m and station reports 500 hpa level

The reporting practice is also considered as bad if the station reports 500 gpm, 1000 gpm, 2000 gpm, 3000 gpm or 4000 gpm level pressure, respectively, and station altitude is more than 800 m different from the reported level.

For temp geopotentials the reporting practice is considered as bad if:

- station altitude is above 800 m and station reports 1000 hPa level
- station altitude is above 2300 m and station reports 850 hPa level
- station altitude is above 3700 m and station reports 700 hPa level

A.2 Revised background quality control for selected observations

The background quality-control rejection limits are applied more strictly for some observation types than stated in Table 10.1. The special cases are the following ones.

- AIREP wind observations with zero wind speed are rejected if the background wind exceeds 5 ms^{-1}.
- For AIREP and DRIBU wind observations the rejection limit is multiplied by 0.5, and for pilot wind by 0.8.
- For SATOB wind observations the rejection limit is multiplied by 0.1, except below 700 hPa level where it is multiplied by 0.2.
- No background quality control is applied for SCAT winds.
- For DRIBU surface pressure observations the rejection limit is multiplied by 0.9, and for paob surface pressure by 0.7.
- For AIREP temperature observations the rejection limit is multiplied by 1.6.

A.3 Use of atmospheric motion winds

This appendix describes those parts of the ECMWF assimilation system which involves some special code for the AMW case, i.e. the data selection and the FG quality check. It refers to the operational status as from December 1996. A thinning procedure was introduced for high-density winds in Spring 1998.

A.3.1 Data selection

There are several model independent checks which AMW data have to pass in order to be considered for the assimilation process:.

Check on longitude/latitude

- AMW must be within a circle of 55° from the sub-satellite point

Check on levels depending on the computational method

- WW CMW and WVMW must be above 400 hPa
- VIS CMW must be below 700 hPa
- IR CMW can be used at all levels.

Check on land/sea

- All AMW over sea are used.
- AMW over land is not used north of 20°N.
- For Meteosat (0° mission) instead of 20°N this threshold is 35°N to allow usage of AMW over north Africa.
- For Meteost (63° mission) the use of AMW has been extended over Asia if above 500 hPa. This is restricted for longitudes east of 30°E.
- AMW are blacklisted over the Himalayas as a precautionary measure.
- AMW over land south of 20°N (35°N for Meteosat) is used if above 500 hPa.

Check on satellite (35°N for Meteosat) is used if above 500 hPa.

This is a temporary selection on certain channels or satellites. At present channels and satellite used are:

- METEOSAT cloud tracked winds with 90 min temporal sampling
- METEOSAT IR (not at medium level), VIS, WV
- METEOSAT HVIS, also at asynoptic times, only if $QI_2 \equiv 0$ (Automatic Quality Control $\equiv$ PASSED)
- GOES IR & WV (NOT at asynoptic times)
- GMS IR & VIS

A.3.2 Background quality check

The background quality check is based on a comparison of the AMW deviation from the background. Observed wind components are checked together. The AMW is flagged with $j = 1$ or 2 or 3 if this deviation squared is greater than a predetermined multiple ERRLIM * ZREJMOD of its estimated variance, as given by the following expression:

if [D2 > (sfg 2 + sobs 2) * ERRLIMj * ZREJMOD] then flag = j where D 2 = 1/2 (Du2 + D v2) with Du, Dv wind component deviations from background; sfg std of the background wind component error (mean for u and v); sobs std of the observation wind component error, 2 m s^{-1} for levels below 700 hPa included, 3.5 m s^{-1} at 500 hPa, 4.3 m s^{-1} at 400 hPa and 5 m s^{-1} for all levels above; ERRLIMj is 8 for $j = 1$, 18 for $j = 2$ and 20 for $j = 3$. The value of ZREJMOD depends on the level of AMW and normally its value is:

- ZREJMOD = 0.2 for low level
- ZREJMOD = 0.1 for all others levels

A special check or asymmetric check is applied when the observed speed is more than 4 m s^{-1} slower than the background speed SPDfg. This check has a more restrictive rejection limit:

- ZREJMOD = 0.15 at low level
- ZREJMOD = 0.07 in the tropics
- ZREJMOD = 0.075 − 0.00125 * SPDfg all others
- ZREJMOD = 0.0 if SPDfg > 60 m s^{-1} (observation gets always flag $j = 3$)

When the data is passed to the following variational quality control its probability of being used depend on the flag j. With flag $j = 1$ the data will be assimilated, with flag $j = 2$ it will be given an intermediate probability and might be used or not and finally the analysis will reject all data with $j = 3$.

Chapter 11
Analysis of snow

Table of contents

Snow depth is a model prognostic variable that needs to be analysed. Its analysis is performed in a module that is currently separated from the analysis of the atmosphere and of the soil wetness. This module includes also the sea-surface temperature, sea-ice fraction and screen-level temperature and relative humidity.

11.1 ORGANIZATION

The snow analysis is a 3-D sequential analysis performed every 6 hours using a successive correction method. The snow-depth background S^{b} (units: m) is estimated from the short-range forecast of snow water equivalent $W_{\mathrm{s}}^{\mathrm{b}}$ (units: m of water equivalent) and snow density $\rho_{\mathrm{s}}^{\mathrm{b}}$ (units: kg m^{-3}). It is given by

$$S^{\mathrm{b}} = \frac{1000 \times W_{\mathrm{s}}^{\mathrm{b}}}{\rho_{\mathrm{s}}^{\mathrm{b}}} \tag{11.1}$$

The snow analysis S^{a} is performed using snow-depth observations, the snow-depth background field, and NOAA/ NESDIS snow extent. If snow-depth observations are not available, the snow accumulation/melting is simulated from the model six-hour forecast. The use of the satellite derived snow extent is optional. As an alternative, snow climate can be used to ensure the stability of the scheme and to give a seasonal snow trend in areas without any observations.

11.2 SNOW-DEPTH ANALYSIS

The snow analysis is a two-step algorithm (Drusch *et al.*, 2004). In the first step, the background field S^{b} (as defined above) is compared with the NOAA/NESDIS snow extent product. Grid boxes, which are snow free in the first guess but snow covered in the satellite derived product are updated with a constant snow depth of 10 cm. In the second step, the actual Cressman analysis is performed based on observations S^{o} from SYNOP reports and snow free satellite observations, which enter the analysis with a snow depth of 0 cm. The actual interpolation is given through:

$$S^{\mathrm{a}} = S^{\mathrm{b}} + \frac{\sum_{n=1}^{N} w_n (S_n^{\mathrm{o}} - S^{\mathrm{b}'})}{\sum_{n=1}^{N} w_n} \tag{11.2}$$

The weight function w_n is the product of functions of the horizontal distance r and vertical displacement h (model minus obs height) between the observation and analysis points so that

$$w = H(r)v(h) \tag{11.3}$$

where

$$H(r) = \max\left(\frac{r_{\max}^2 - r^2}{r_{\max}^2 + r^2}, 0\right) \tag{11.4}$$

and

$$v(h) = 1 \quad \text{if } 0 < h$$
$$v(h) = \frac{h_{\max}^2 - h^2}{h_{\max}^2 + h^2} \quad \text{if } -h_{\max} < h < 0$$
$$v(h) = 0 \quad \text{if } h < -h_{\max}$$

The snow depth is preserved when the model height is above the observing station, but it is severely reduced below. The observation height for the satellite data is obtained from the model orography field. The influence distances are set to $r_{\max} = 250$ km and $h_{\max} = 300$ m.

In addition to the preliminary quality control in the observation data base, the following checks are applied for each grid point.

(i) If $T_{2m}^{\mathrm{b}} < 8°\mathrm{C}$ only snow depth observations below 140 cm are accepted.
(ii) This limit is reduced to 70 cm if $T_{2\mathrm{m}}^{\mathrm{b}} > 8°\mathrm{C}$.
(iii) Snow-depth observations are rejected if they differ by more than 50 cm from the background.
(iv) When only one snow-depth observation is available within the influence radius $r_{\max}$, the snow depth increments are set to zero.
(v) Snow-depth analysis is limited to 140 cm.
(vi) Snow-depth increments are set to zero when larger than $(160 - 16T_{2\mathrm{m}}^{\mathrm{b}})$ mm (where $T_{2\mathrm{m}}^{\mathrm{b}}$ is expressed in Celsius).
(vii) Snow-depth analysis is set to zero if below 0.04 cm.
(viii) If there is no snow in the background and in more than half of the observations within a circle of radius $r_{\max}$, the snow-depth increment is kept to zero.

It is expected that the satellite derived snow extent can be used in the analysis to replace the role of snow depth climatology in correcting for the model bias. However, there is the option to weight the analysis of snow depth with the climatological value S^{clim} so that the final analysis is provided by

$$S^{\mathrm{a}} = (1 - \alpha)S^{\mathrm{a}} + \alpha S^{\mathrm{clim}} \tag{11.5}$$

The relaxation coefficient α can be changed through the namelist (see Section 11.3). Its default value is set to 0.02, which corresponds to a time scale of 12.5 days at six-hourly cycling.

The final snow water equivalent product $W_{\mathrm{s}}^{\mathrm{a}}$ is then calculated using

$$W_{\mathrm{s}}^{\mathrm{a}} = \frac{\rho_{\mathrm{s}}^{\mathrm{b}} \times S^{\mathrm{a}}}{1000} \tag{11.6}$$

The snow density is unchanged in the analysis process: $\rho_{\mathrm{s}}^{\mathrm{a}} = \rho_{\mathrm{s}}^{\mathrm{b}}$.

Areas with permanent snow and ice (defined using the Global Land Cover Characterization product) are set to an arbitrary high value at each analysis cycle ($W_{\mathrm{s}}^{\mathrm{a}} = 10$ m).

The snow temperature analysis is performed within the soil analysis as described in Section 12.3.

11.3 TECHNICAL ASPECTS

The snow analysis software is implemented as a branch of the more comprehensive surface and screen-level analysis (SSA) package. The other branches currently include two-metre temperature and relative humidity analysis, and also sea surface temperature and sea-ice fraction analyses. The program organization when performing snow analysis is roughly as follows.

- SSA
  - CONTROL_SSA

- INISNW
  - SCAN_DDR
  - COORDINATES
  - GETFIELDS
  - sub_prep_ues.F90
  - SCAN_CMA
    - SCAN_OBS
      - LAND_OBS
      - INITIAL_REJECTION
      - REDUNDANT_OBS
- SNOW_ANALYSIS
  - CRES_FILL
    - sub_prep_nes.F90
  - SUCSNW
  - SCAN_OBS
    - FG2OBS
  - SUCSNW
  - SNOW_FG
- FDB_OUTPUT
- PRINT_SUMMARY
- PLOTDATA
- FEEDBACK

The main program SSA calls CONTROL_SSA where most of the setup and namelist handling are done. Routine INISNW performs initialization of the actual snow analysis by sensing the size of the observation array file (CMA-file) in SCAN_DDR and generating latitudinal coordinates that stem from the model resolution in concern and zeros of the Bessel function.

After this, all input fields are read into memory in GETFIELDS. They consist of the snow water equivalent and snow density from the first-guess (6-hour forecast), 2 m temperature first guess, snow-depth climate (varies monthly with a linear temporal interpolation), land/sea mask and finally the orography in a form of the geopotential.

In sub_prep_nes.F90 the satellite data, which contain no snow are retrieved from buffr.

In SCAN_CMA observations are read into memory and a quick validity check of the non-applicable observations for this analysis is performed. Furthermore, the land/sea mask is calculated in LAND_OBS for the retained snow depth observation points.

Additional screening is done in INITIAL_REJECTION and in REDUNDANT_OBS. The former one sets up an internal table where all the observations which survived from the quick screening are placed with a minimum context information. This routine rejects some of the observations entered into the table due to inconsistencies.

The routine REDUNDANT_OBS removes time duplicates and retains the observations of the station in concern with the closest (and the most recent) to the analysis time. Since only synoptic observations are considered, slowly moving platform handling present in the REDUNDANT_OBS is not applicable to the snow analysis.

The actual snow analysis is performed under SNOW_ANALYSIS. In a first step, the first guess field is updated with the satellite observations (routine CRES_FILL). The satellite observations, which contain snow are retrieved from buffr (routine SUB_PREP_NES). The analysis technique is Cressman's successive correction method (routine SUCSNW). The structure functions are set to be separable in horizontal and vertical directions. A special mountain region handling is performed, depending whether the datum or grid points is in the valley or at high altitudes, as explained before.

The snow-depth background (i.e. first guess) field is constructed from the model first-guess snow water equivalent and snow density. Once the snow-depth first guess field is present, it is used to calculate the first

guess departure at snow-depth observation points. This increment is finally added to the snow depth fields at grid points producing the final snow depth output field, which is output in routine FDB_OUTPUT.

The accuracy of the analysis is estimated in PRINT_SUMMARY where some important statistics are summarized. The internal observation table can be printed if requested from PLOTDATA and an updated observation file for feedback purposes can be created in routine FEEDBACK.

The main logicals of the namelist NAMSSA are as follows.

(i) L_SNOW_ANALYSIS: When set to .TRUE., the snow analysis is performed.
(ii) L_SNOW_DEPTH_ANA: When set to .TRUE., the snow analysis is performed in snow depth (in opposition to snow water equivalent assuming a constant value of 250 kg m^{-2} for observed snow density).
(iii) L_USE_SNOW_CLIMATE: When set to .TRUE., a relaxation of the snow analysis towards a monthly climatology is performed with a time scale of 12.5 days (the relaxation coefficient is passed through NAMSSA).
(iv) L_USE_FG_FIELD: When set to .TRUE. the snow analysis is set to the first-guess value (no use of observations) and there is no relaxation to climatology.
(v) L_USE_SCOVER_NESDIS: When set to .TRUE., the NOAA/NESDIS satellite product is used.

Chapter 12
Land-surface analysis

Table of contents

12.1 INTRODUCTION

Soil temperature and soil water content are prognostic variables of the forecasting system and, as a consequence, they need to be initialised at each analysis cycle. Currently the land surface analysis is performed every 6 hours and is decoupled from the atmospheric analysis. The absence of routine observations on soil moisture and soil temperature requires to use proxy data. The ECMWF soil analysis relies on SYNOP temperature and relative humidity at screen-level (2 m) available on the GTS (around 12,000 reports over the globe are provided every 6 hours). Firstly, a screen-level analysis is performed for temperature and humidity. Secondly, the screen-level analysis increments are used as inputs to perform the analysis in the soil.

12.2 SCREEN-LEVEL ANALYSIS

12.2.1 Methodology

Two independent analyses are performed for 2 m temperature and 2 m relative humidity. The method used is a two-dimensional univariate statistical interpolation. In a first step, the background field (6 h or 12 h forecast) is interpolated horizontally to the observation locations using a bilinear interpolation scheme and background increments ΔX_i are estimated at each observation location i.

The analysis increments ΔX_j^a at each model grid-point j are then expressed as a linear combination of the first-guess increments (up to N values) given by

$$\Delta X_j^{\mathrm{a}} = \sum_{i=1}^{N} W_i \times \Delta X_i \tag{12.1}$$

where W_i are optimum weights given (in matrix form) by

$$(\mathbf{B} + \mathbf{O})\mathbf{W} = \mathbf{b} \tag{12.2}$$

The column vector $\mathbf{b}$ (dimension N) represents the background error covariance between the observation i and the model grid-point j. The $(N \times N)$ matrix $\mathbf{B}$ describes the error covariances of background fields between pairs of observations. The horizontal correlation coefficients (structure functions) of $\mathbf{b}$ and $\mathbf{B}$ are assumed to have the form

$$\mu(i,j) = \exp\left(-\frac{1}{2}\left[\frac{r_{ij}}{d}\right]^2\right) \tag{12.3}$$

where r_{ij} is the horizontal separation between points i and j and d the e-folding distance taken to 300 km (hard coded in subroutine OIINC).

Therefore

$$B(i,j) = \sigma_b^2 \times \mu(i,j) \tag{12.4}$$

with σ_b the standard deviation of background errors.

The covariance matrix of observation errors $\mathbf{O}$ is set to $\sigma_o^2 \times \mathbf{I}$ where σ_o is the standard deviation of observation errors and $\mathbf{I}$ the identity matrix.

The standard deviations of background and observation errors are set respectively to 1.5 K and 2 K for temperature and 5% and 10% for relative humidity. The number of observations closest to a given grid point that are considered to solve (12.1) is $N = 50$ (scanned within a radius of 1000 km). The analysis is performed over land and ocean but only land (ocean) observations are used for model land (ocean) grid points.

12.2.2 Quality controls

Gross quality checks are first applied to the observations such as $RH \in [2, 100]$ and $T > T^d$ where T^d is the dewpoint temperature. Redundant observations are also removed by keeping only the closest (and more recent) to the analysis time.

Observation points that differ by more than 300 m from the model orography are rejected.

For each datum a check is applied based on statistical interpolation methodology. An observation is rejected if it satisfies :

$$|\Delta X_i| > \gamma \sqrt{\sigma_o^2 + \sigma_b^2} \tag{12.5}$$

where γ has been set to 3, both for temperature and humidity analyses.

The number of used observations every 6 hours varies between 4,000 and 6,000 corresponding to around 40% of the available observations.

The final relative humidity analysis is bounded between 2% and 100%. The final MARS archived product is dew-point temperature that uses the 2 m temperature analysis T_a to perform the conversion so that

$$T^d = \frac{17.502 \times 273.16 - 32.19 \times \Psi}{17.05 - \Psi} \tag{12.6}$$

with

$$\Psi = \log(RH_a) + 17.502 \times \frac{T_a - 273.16}{T_a - 32.19} \tag{12.7}$$

12.2.3 Technical aspects

The technical aspects are similar to the snow analysis (see Chapter 11) expect for the computation of the analysis increments obtained from the subroutine OIUPD instead of SUCSNW (Cressman interpolation).

Subroutine OISET selects the N closest observations from a given grid-point.

Subroutine OIINC provides the analysis increments from (12.1)and (12.2), by first computing $\mathbf{q} = (\mathbf{B} + \mathbf{O})^{-1}\Delta\mathbf{X}$ (in subroutine EQUSOLVE – inversion of a linear system) which does not depend upon the position of the analysis gridpoint and then estimating $\mathbf{b}^T\mathbf{q}$ (in subroutine DOT_PRODUCT).

Most of the control parameters of the screen-level analysis are defined in the namelist NAMSSA.

(i) C_SSA_TYPE: 't2m' for temperature analysis and 'rh2m' for relative humidity analysis.
(ii) L_OI : .TRUE. for statistical interpolation and .FALSE. for Cressman interpolation.
(iii) N_OISET: number of observations (parameter N).
(iv) SIGMAB: standard deviation of background error (parameter σ_b).
(v) SIGMAO: standard deviation of observation error (parameter σ_o).
(vi) TOL_RH: Tolerance criteria for RH observations (parameter γ in (12.5)).
(vii) TOL_T: Tolerance criteria for T observations (parameter γ in (12.5)).
(viii) SCAN_RAD_2M(1): Scanning radius for available observations (set to 1000 km).

12.3 SOIL ANALYSIS

The soil analysis scheme is based on an "local" optimum interpolation technique as described in Mahfouf (1991) and Douville *et al.* (2001). The analysis increments from the screen-level analysis are used to produce increments for the water content in the first three soil layers (corresponding to the root zone) given by

$$\Delta\theta = a \times (T_{\rm a} - T_{\rm b}) + b \times \left[100\frac{e_{\rm s}(T_{\rm a}^{\rm d}) - e_{\rm s}(T_{\rm b}^{\rm d})}{e_{\rm s}(T_{\rm a})}\right] \tag{12.8}$$

and for the first layer soil temperature and snow temperature

$$\Delta T = c \times (T_{\rm a} - T_{\rm b}) \tag{12.9}$$

with $T_{\rm a}$ and $T_{\rm b}$ the analysed and model first guess temperatures, respectively. The coefficients a and b are defined as the product of optimum coefficients α and β minimising the variance of analysis error and of empirical functions F_1, F_2, and F_3 reducing the size of the optimum coefficients when the coupling between the soil and the lower boundary layer is weak.

$$\alpha = \frac{\sigma_\theta}{\phi\sigma_{\rm b}^T}\left\{\left[1 + \left(\frac{\sigma_{\rm a}^{RH}}{\sigma_{\rm b}^{RH}}\right)^2\right]\rho_{T\theta} - \rho_{RHT}\rho_{RH\theta}\right\} \tag{12.10}$$

and

$$\beta = \frac{\sigma_\theta}{\phi\sigma_{\rm b}^{RH}}\left\{\left[1 + \left(\frac{\sigma_{\rm a}^{RH}}{\sigma_{\rm b}^{RH}}\right)^2\right]\rho_{RH\theta} - \rho_{RHT}\rho_{T\theta}\right\} \tag{12.11}$$

with

$$\phi = \left[1 + \left(\frac{\sigma_{\rm a}^{T}}{\sigma_{\rm b}^{T}}\right)^2\right]\left[1 + \left(\frac{\sigma_{\rm a}^{RH}}{\sigma_{\rm b}^{RH}}\right)^2\right] - \rho_{RHT}^2 \tag{12.12}$$

where ρ_{xy} represents the correlation of background errors between parameters x and y.

The statistics of background errors have been obtained from a series of Monte-Carlo experiments with a single-column version of the atmospheric model where initial conditions for soil moisture have been perturbed randomly. They were obtained for a clear-sky situation with strong solar insolation. Empirical functions are aimed to reduce soil increments when atmospheric forecast errors contain less information about soil moisture. To obtain negligible soil-moisture corrections during the night and in winter, F_1 is a function of the cosine of the mean solar zenith angle $\mu_{\rm M}$, averaged over the 6 h prior to the analysis time given by

$$F_1 = \frac{1}{2}\{1 + \tanh[\lambda(\mu_{\rm M} - 0.5)]\} \quad \lambda = 7 \tag{12.13}$$

The optimum coefficients are also reduced when the radiative forcing at the surface is weak (cloudy or rainy situations). For this purpose, the atmospheric transmittance $\tau_{\rm r}$ is computed from the mean downward surface solar radiation forecasted during the previous 6 hours $\langle R_{\rm g}\rangle$ as

$$\tau_{\rm r} = \left(\frac{\langle R_{\rm g}\rangle}{S_{\rm o}\mu_{\rm M}}\right)^{\mu_{\rm M}} \tag{12.14}$$

where $S_{\rm o}$ is the solar constant.

The empirical function F_2 is expressed as

$$F_2 = \begin{cases} 0 & \tau_{\rm r} < \tau_{\rm r\,min} \\ \dfrac{\tau_{\rm r} - \tau_{\rm r\,min}}{\tau_{\rm r\,max} - \tau_{\rm r\,min}} & \tau_{\rm r\,min} < \tau_{\rm r} < \tau_{\rm r\,max} \\ 1 & \tau_{\rm r} > \tau_{\rm r\,max} \end{cases} \tag{12.15}$$

with $< \tau_{\rm r\,min} = 0.2$ and $< \tau_{\rm r\,max} = 0.9$.

The empirical function F_3 reduces soil moisture increments over mountainous areas so that

$$F_3 = \begin{cases} 0 & Z > Z_{\rm max} \\ \left(\dfrac{Z - Z_{\rm max}}{Z_{\rm min} - Z_{\rm max}}\right)^2 & Z_{\rm min} < Z < Z_{\rm max} \\ 1 & Z < Z_{\rm min} \end{cases} \tag{12.16}$$

Table 12.1 *Statistics of background errors for soil moisture derived from Monte-Carlo experiments.*

Coefficient	Value
$\rho_{T\theta 1}$	-0.82
$\rho_{T\theta 2}$	-0.92
$\rho_{T\theta 3}$	-0.90
$\rho_{RH\theta 1}$	0.83
$\rho_{RH\theta 2}$	0.93
$\rho_{RH\theta 3}$	0.91
σ_{b}^{T}	1.25 K
σ_{b}^{RH}	9.5%
ρ_{RHT}	-0.99

where Z is the model orography, $Z_{\min} = 500$ m and $Z_{\max} = 3000$ m.

Furthermore, soil moisture increments are set to zero if one of the following conditions is fulfilled:

(i) The last 6 h precipitation exceeds 0.6 mm.
(ii) The instantaneous wind speed exceeds 10 ms^{-1}.
(iii) The air temperature is below freezing.
(iv) There is snow on the ground.

To reduce soil moisture increments over bare soil surfaces, the standard deviations and the correlations coefficients are also weighted by the vegetation fraction $C_{\mathrm{v}} = c_{\mathrm{L}} + c_{\mathrm{H}}$, where low and high vegetation cover are defined in Chapter 7 of Part IV.

The statistics of forecast errors necessary to compute the optimum coefficients are given in Table 12.1.

The correlations have been produced from the Monte-Carlo experiments. The standard deviation of background error for soil moisture σ_θ is set to 0.01 $\mathrm{m}^3\ \mathrm{m}^{-3}$ on the basis of ECMWF forecasts differences between day 1 and day 2 of the net surface water budget (precipitation minus evaporation minus runoff).

The standard deviation of analysis error σ_{a} is given by the screen-level analysis from

$$\frac{1}{\sigma_{\mathrm{a}}^2} = \frac{1}{\sigma_{\mathrm{b}}^2} + \frac{1}{\sigma_{\mathrm{o}}^2} \tag{12.17}$$

From the values chosen for the screen-level analysis $\sigma_{\mathrm{a}}^{T} = 1.2$ K and $\sigma_{\mathrm{a}}^{RH} = 4.47\%$.

Soil moisture increments $\Delta\theta$ are such that they keep soil moisture within the wilting point θ_{pwp} and the field capacity θ_{cap} values. That is

(i) If $\theta_{\mathrm{b}} < \theta_{\mathrm{cap}}$ then $\theta_{\mathrm{a}} = \min(\theta_{\mathrm{cap}}, \theta_{\mathrm{b}} + \Delta\theta)$
(ii) If $\theta_{\mathrm{b}} > \theta_{\mathrm{pwp}}$ then $\theta_{\mathrm{a}} = \max(\theta_{\mathrm{pwp}}, \theta_{\mathrm{b}} + \Delta\theta)$

Finally the coefficients providing the analysis increments are

$$\begin{aligned} a &= C_{\mathrm{v}} \times \alpha \times F_1 F_2 F_3 \\ b &= C_{\mathrm{v}} \times \beta \times F_1 F_2 F_3 \end{aligned} \tag{12.18}$$

and

$$C = (1 - F_1) F_3 \tag{12.19}$$

The coefficient c is such that soil and snow temperatures are more effective during night and in winter, when the temperature errors are less likely to be related to soil moisture. This way, 2 m temperature errors are not used to correct soil moisture and soil temperature at the same time.

In the 12 h 4D-Var configuration, the soil analysis is performed twice during the assimilation window and the sum of the increments is added to the background values at analysis time.

Chapter 13

Analysis of sea-ice concentration and sea surface temperature

Table of contents

13.1 INTRODUCTION

The analyses of sea-ice concentration (CI) and sea surface temperature (SST) are based on daily global 0.5° datasets provided by NCEP. The sea-ice product from NCEP is derived from SSM/I (Special Sensor Microwave/Imager) observations. The SST product is an Optimal Interpolation (OI) analysis based on ship, buoy, and satellite measurements. At ECMWF, the NCEP fields are re-sampled to the reduced Gaussian grid used in the IFS, and are quality checked.

CI and SST are analysed six-hourly before the atmospheric variational analysis. Since sea-ice concentration and SST are kept constant during the ten-day forecast and the NCEP fields enter the analysis once a day, the analyses are updated once a day. In the current set up, the CI analysis is done before the SST analysis. The SST field from the previous analysis is used as a quality check in the ice analysis. The CI analysis is then used to determine the SST for grid boxes containing sea ice. Both analyses are performed for model sea grid boxes (i.e. those, which contain less than 50% land). Only the Great Lakes and the Caspian Sea are covered in the NCEP datasets.

13.2 SEA-ICE CONCENTRATION

ECMWF's sea-ice concentration analysis is based on the corresponding daily NCEP product as described by Grumbine (1996). The primary source of information is 19 GHz and 37 GHz SSM/I antenna temperatures. This document describes the processing at NCEP including the ice algorithm and the re-sampling to the reduced Gaussian grid used in the IFS.

13.2.1 NCEP dataset

This section describes the final quality controlled NCEP dataset, which is the input to the IFS, following Grumbine (1996). The sea-ice concentration algorithm, which generates total, first-year, and multiyear concentration from one set of SSM/I brightness temperatures is introduced in Subsection 13.2.1.

(*a*) *Brightness-temperature processing*

Brightness-temperature grids are produced using a 12-hour window, before and after 0000 UTC of the current day. The data extraction program is run at 1230 or 1330 UTC (8:30 Eastern Local Time) and, in general, 12 orbits out of 14 possible orbits are included in the analysis. Brightness temperatures

are then calculated from the antenna temperatures included in the Satellite Data Record (SDR) following Hollinger *et al.* (1987). Quality checks are applied at two stages: (i) Antenna temperature scans are removed if the corresponding surface types (encoded into an 8 bit byte) do not fall in the range of 0 to 7. (ii) Brightness-temperature checks are applied using thresholds established by Gloersen *et al.* (1994).

(*b*) *Gridding and computation of sea-ice concentration*

Quality-checked antenna temperatures are remapped from scan points onto a 25.4 km, true at 60°, polar stereographic grid oriented to 80° W in the northern hemisphere, and 100° E in the southern hemisphere. The northern hemisphere grid is 385 × 465 pixels with the pole at (191, 231). The southern hemisphere grid is 345 × 355 pixels with the pole at (151, 181).

After the antenna temperatures are converted to brightness temperatures, sea-ice concentrations are calculated using the algorithm described by (Cavalieri *et al.*, 1991). The algorithm calculates the total (C), first-year ($C_{\rm F}$), and multiyear ($C_{\rm M}$) sea-ice concentrations from the microwave polarization (PR) at 19 GHz ($T_{19H/V}$) and the spectral gradient ratio (G_R), which uses the vertically polarized 37 GHz and 19 GHz brightness temperatures T_{37V} and T_{19V}, respectively (Cavalieri *et al.*, 1991). Therefore

$$PR = (T_{19\rm v} - T_{19\rm H})/(T_{19\rm v} + T_{19\rm H}) \tag{13.1}$$

$$GR = (T_{37\rm v} - T_{19\rm v})/(T_{37\rm v} + T_{19\rm v}) \tag{13.2}$$

C, $C_{\rm F}$, and $C_{\rm M}$ are then obtained through

$$C_{\rm F} = (a_0 + a_1 PR + a_2 GR + a_3 PR \times GR)/D \tag{13.3}$$

$$C_{\rm M} = (b_0 + b_1 PR + b_2 GR + b_3 PR \times GR)/D \tag{13.4}$$

$$C = C_{\rm F} + C_{\rm M} \tag{13.5}$$

with

$$D = c_0 + c_1 PR + c_2 GR + c_3 PR \times GR \tag{13.6}$$

The coefficients a_i, b_i, and c_i $(i = 0, 3)$ are functions of a set of nine brightness temperatures. These brightness temperatures referred to as algorithm tie points are observed SSM/I radiances of known ice-free ocean, first-year sea ice, and multi-year ice for each of the three SSM/I channels.

The weather filter proposed by Gloersen and Cavalieri (1986) is extended to adjacent pairs of points to take the coarse resolution of the 19 GHz observations, which does not match the polar stereographic grid resolution, into account. Since the satellite does not observe the poles, the disk of missing data on the northern hemisphere is filled according to a Laplacean with boundary values specified by the observations. This method gives the minimum gradient fill pattern. In addition, the Laplacean of ice concentration is small in regions away from the ice edge, which the polar data gap always is.

Values for sea-ice concentration vary between 0% and 128%. Above 128, there are flags denoting different data conditions: Land (157), Bad Data (166), Weather (177), Coast (195), and No Data (224).

(*c*) *Final filtering and gridding for NWP*

SST is used as a final filter. If the analysed SST is greater than 2°C, sea-ice concentration is set to 0. As stated by Grumbine (1996), this filter is good as long as low concentrations are not being considered important. For NWP applications the polar stereographic product is sampled to 0.5° × 0.5° resolution using the nearest-neighbour technique.

13.2.2 ECMWF re-sampling to model grid

At ECMWF, the 0.5° × 0.5° dataset ($CI^{\rm NCEP}$) is currently being used operationally. A bi-linear interpolation is applied to transform the NCEP data from the regular latitude/longitude grid to the reduced Gaussian model grid. Before May 2001, the CI product in polar stereographic projection was used. In this case, a Cressman spatial interpolation is performed to transform the data to the reduced

Gaussian grid using

$$CI^{\mathrm{a}} = CI^{\mathrm{b}} + \frac{\sum_{n=1}^{N} w_n (CI_n^{\mathrm{NCEP}} - CI^{\mathrm{b}'})}{\sum_{n=1}^{N} w_n} \tag{13.7}$$

with a maximum radius $r_{\max}$ of 120 km. CI^{b} and $CI^{\mathrm{b}'}$ are the first guess (ice concentration from the previous day) at the model grid point and at the observation, respectively. The weighting function is given by

$$w(r) = \frac{r_{\max} - r}{r_{\max} + r} \tag{13.8}$$

In the next step, a high resolution CI analysis product from the Swedish Meteorological and Hydrological Institute (SMHI) is incorporated to replace the NCEP data set over the Baltic Sea area as described in Drusch (2006).

Finally, the following quality checks are applied.

(i) Sea-ice concentration below 20% is set to 0%.
(ii) Sea-ice concentration greater 100% is set to 100%.
(iii) If the SST is higher than 1°C sea-ice concentration is set to 0%.
(iv) Grid boxes north of 82.5°N are set to 100%.

If the total change in global ice concentration exceeds 0.2%, the CI analysis is considered to be wrong and aborted.

13.3 SST ANALYSIS

For the operational ten-day forecasts, ECMWF uses the NCEP daily real-time global SST product (RTG_SST) at 0.5° resolution. For the seasonal forecasts the weekly the 1.0° product ('Reynolds SST') is temporally interpolated to daily values.

13.3.1 Daily NCEP dataset

The information provided in this section was extracted from http://polar.wwb.noaa.gov.sst. The daily SST product has been developed at NCEP/MMAB (Marine Modelling and Analysis Branch). It was implemented in the production suite 30 January, 2001.

A two-dimensional variational interpolation analysis of the most recent 24-hour buoy and ship data, satellite retrieved SST data, and SSTs derived from satellite observed sea-ice coverage is used to generate the 0.5° dataset. The algorithm employs the following data-handling and analysis techniques.

(i) Satellite retrieved SST values are averaged within 0.5° grid boxes with day and night 'superobs' created separately for each satellite.
(ii) Bias calculation and removal, for satellite retrieved SST, is the technique employed in the seven-day Reynolds–Smith climatological analysis.
(iii) SST report from individual ships and buoys are separately averaged within grid boxes.
(iv) The first guess is the prior (un-smoothed) analysis with one-day's climate adjustment added.
(v) Late arriving data, which did not make it into the previous SST analysis, are accepted if they are less than 36 hours old.
(vi) Where ice cover exceeds 50%, surface temperature is calculated using salinity climatology in Millero's formula for the freezing point of salt water:

$$t(S) = -0.0575S + 0.0017S^{3/2} - 0.0002S^2$$

with S in psu.
(vii) An inhomogeneous correlation-scale-parameter l, for the correlation function $\exp(-d^2/l^2)$ is calculated from a climatological temperature gradient, as

$$l = \min(450, \max(2.25/|\nabla T|, 100))$$

with d and l in km and ∇T in °C/km.

The RTG_SST analysis is done over all ocean areas, the Great Lakes, and the Caspian Sea. The spatial grid (720×360) is defined with indices starting just east of the Greenwich Meridian and near the South Pole.

13.3.2 ECMWF re-sampling to model grid

The interpolation from the regular latitude/longitude grid to the reduced Gaussian model grid is done by bi-linear interpolation. Ice-free grid boxes are up dated with the latest SST value retrieved from the archive. For grid boxes characterized by sea-ice concentrations exceeding 20% the SST is set to $-1.7°C$.

13.4 TECHNICAL IMPLEMENTATION

Technically, CI and SST analyses are controlled using the logical expressions *L_sst_analysis*, *lssmidata*, *licenmc*, and *lsstnmc*. If *L_sst_analysis* is .TRUE., the analyses are performed (control_ssa.F90). *lssmidata*, *licenmc*, and *lsstnmc* address the input data handling. In the current set up, *lssmidata* and *lsstnmc* are set to .TRUE. and *licenmc* is set to .FALSE. (setcomssa.F90). Whenever there are no ice data in polar stereographic projection available (which is the case from 5 May 2001 onwards) *lssmidata* is set to .FALSE. and *licenmc* is set to .TRUE. in sstana where the namelist for the analyses is created. The actual CI and SST fields on the regular 0.5° grid are retrieved from the MARS archive (fetchmars). The CI product on polar stereographic projection, which was used before May 2001, is stored in buffr format and can be obtained through the fetchobs script.

The SST and CI analyses software is implemented as a branch of the more comprehensive surface and screen-level analysis (SSA) package. The other branches currently include two-meter temperature and relative humidity analysis, and also snow water equivalent analysis. The main program ssa.F90 calls control_ssa.F90. Within control_ssa.F90 variables are initialized (through setcomssa.F90), the namelist is read, and the ODB and the analyses are initialized (through inisst.F90). Major components of the initialization are the interpolation of the regular latitude/longitude fields to the reduced Gaussian model grid and the conversion from the resulting field to a one-dimensional array. The key components for this process are coordinates.F90, reg_to_gg.F90, reg2gg.F90, getfields.F90 and field_to_array.F90.

The actual CI analysis is then performed in ice_analysis.F90. In case the polar stereographic product is used, the buffr file is processed in getsatbuff.F90 and the interpolation to the reduced Gaussian grid is done in ice_cressman.F90. The first guess CI field is updated through the NCEP interpolated product and the quality checks given in Subsection 13.2.2 are applied. Output statistics are prepared and printed in the corresponding log file (sst.1).

The SST analysis is done within sst_analysis.F90. The first guess field is updated through the interpolated NCEP field. Correction based on the actual CI analysis are applied as outlined in Subsection 13.3.2. The log file for the output is sst.1. The following is the calling tree for the CI/SST analyses.

- CONTROL_SSA
  - SETCOMSSA
  - INISST
    - COORDINATES
    - REG_TO_GG
      - REG2GG
    - GETFIELDS
    - FIELD2ARRAY
  - ICE_ANALYSIS
    - GETSATBUFF
    - ICE_CRESSMAN
      - LATBAND
      - CALC_DISTANCE
      - CORWHD
  - SST_ANALYSIS

- FDB_OUTPUT
- PRINT_SUMMARYsea-ice
- PLOTDATA

Chapter 14
Data flow

Table of contents

14.1 NOTATION

The following environment variables, which are used in the same way in the data assimilation scripts, are referred to in this chapter.

Table 14.1 *Definition of environment variables.*

Variable	Meaning	Default value
${DATA}	Data directory for invariant files	
${GTYPE}	Gaussian grid type	l_2 (ie linear reduced Gaussian grid)
${IFS_CYCLE}	IFS cycle name	CY31R1
${LEVELS}	Number of vertical model levels	91
${MM}	Month	
${RESOL}	Spectral truncation	799
${starttime}	Start of 4D-Var window as yyyymmddhh	
${WDIR}	Work directory (1 for each cycle)	

In this chapter, the notation illustrated in Fig. 14.1 is used in diagrams to distinguish between computation steps and data sets.

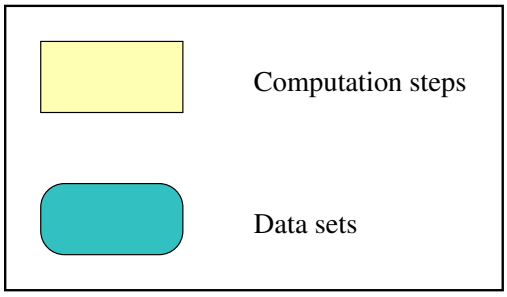

Figure 14.1 *Notation.*

14.2 DATA ASSIMILATION CYCLING

Fig. 14.2 gives an overview of the data flow through the data assimilation system with the operational early-delivery configuration. The 12-hour 4D-Var analyses are run with a delayed cut-off time, in order to use the maximum possible number of observations. The 0000 UTC analysis uses observations from the time window 2101–0900 UTC, while the 1200 UTC analysis uses observations in the window 0901–2100 UTC. The extraction tasks for observations in the periods 2101–0300 UTC and 0301–0900 UTC are run at 1345 and 1400 UTC respectively, while the extraction tasks for the observations in the periods 0901–1500 UTC and 1501–2100 UTC are run at 0145 and 0200 UTC. The 0000 UTC 12-hour 4D-Var analysis generates two sets of analysed fields, at 0000 and 0600 UTC. A separate surface analysis is run every 6 hours. The final analysis is a combination of the fields from 4D-Var and from the surface analysis. The first guess for the 0000 UTC 12-hour 4D-Var analysis is the three-hour forecast from the previous day's 1800 UTC delayed cut-off analysis. The first guess for the 1200 UTC 12-hour 4D-Var delayed cut-off analysis is the three-hour forecast from the 0600 UTC analysis. It is these 12-hour 4D-Var delayed cut-off analyses that propagate information forwards from day to day.

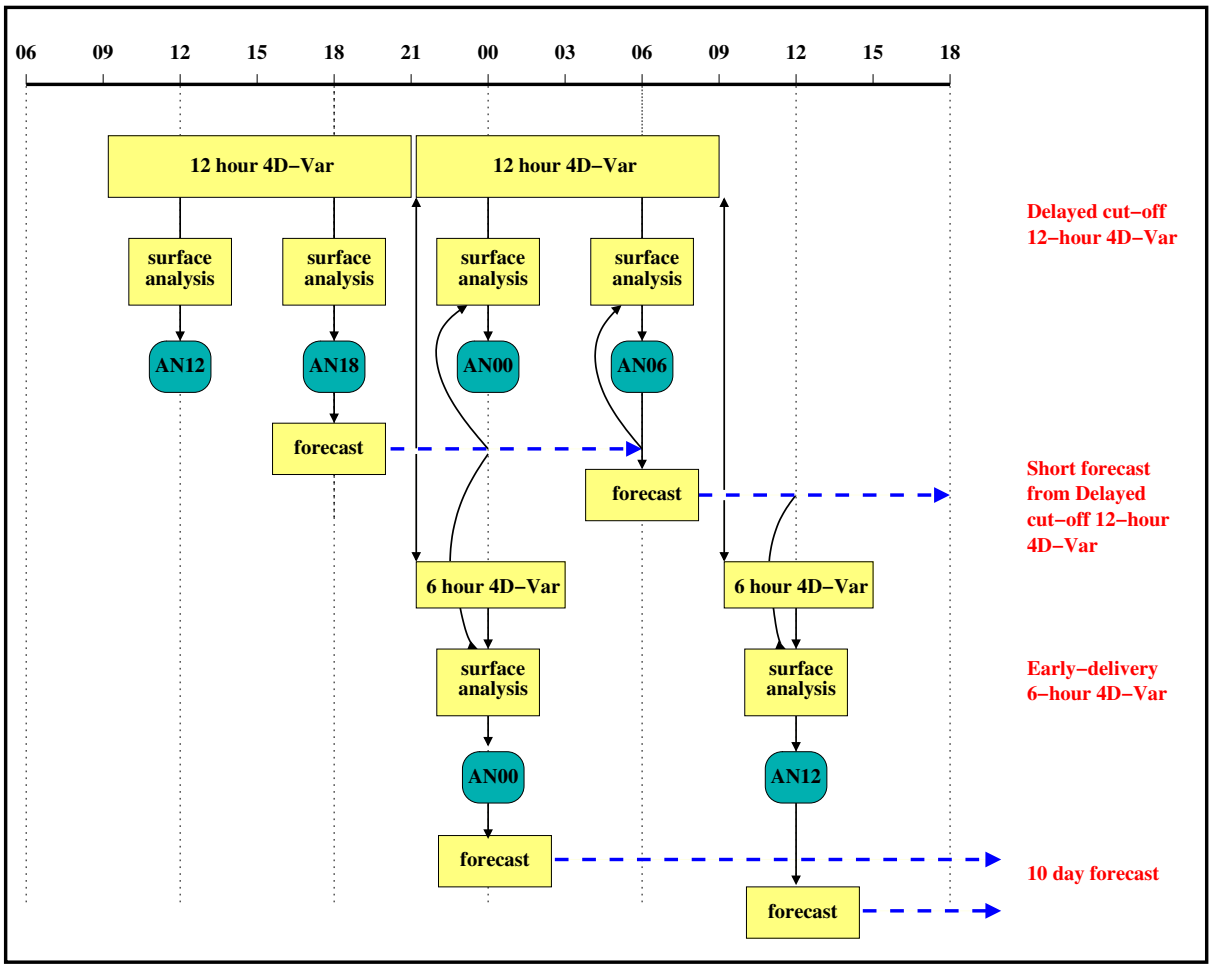

Figure 14.2 *Data assimilation cycling with the Early Delivery configuration.*

The early-delivery analyses do not propagate information from cycle to cycle. Each analysis is reinitialized with the best available model fields from the delayed cut-off assimilation. The 0000 UTC early-delivery analysis is a 6-hour 4D-Var analysis that uses observations in the time window 2101–0300 UTC. The cut-off time is 0400 UTC, and any observations which arrive after this time are not used by the early-delivery analysis. However, if they arrive by 1400 UTC, they can still be used by the delayed cut-off 12-hour 4D-Var 0000 UTC analysis. The first guess for the 0000 UTC early-delivery analysis is the three-hour forecast from the previous day's 1800 UTC delayed cut-off analysis.

The early-delivery 1200 UTC analysis is a 6-hour 4D-Var analysis that uses observations in the time window 0901–1500 UTC, with a cut-off time of 1600 UTC. Its first guess is the three-hour forecast from the 0600 UTC delayed cut-off analysis.

14.3 OVERVIEW OF 4D-VAR DATA FLOW

Fig. 14.3 gives an overview of the data input to and output from 4D-Var. There are three types of input data.

(i) GRIB fields from the Fields Data Base (FDB). GRIB is a World Meteorological Organisation (WMO) standard format for the representation of General Regularly-distributed Information in Binary. The GRIB code is described at http://www.ecmwf.int/products/data/software/grib.html. The background fields and forecast errors are read from the Fields Data Base.
(ii) Observations from the Observation Data Base (ODB).
(iii) Other data files.

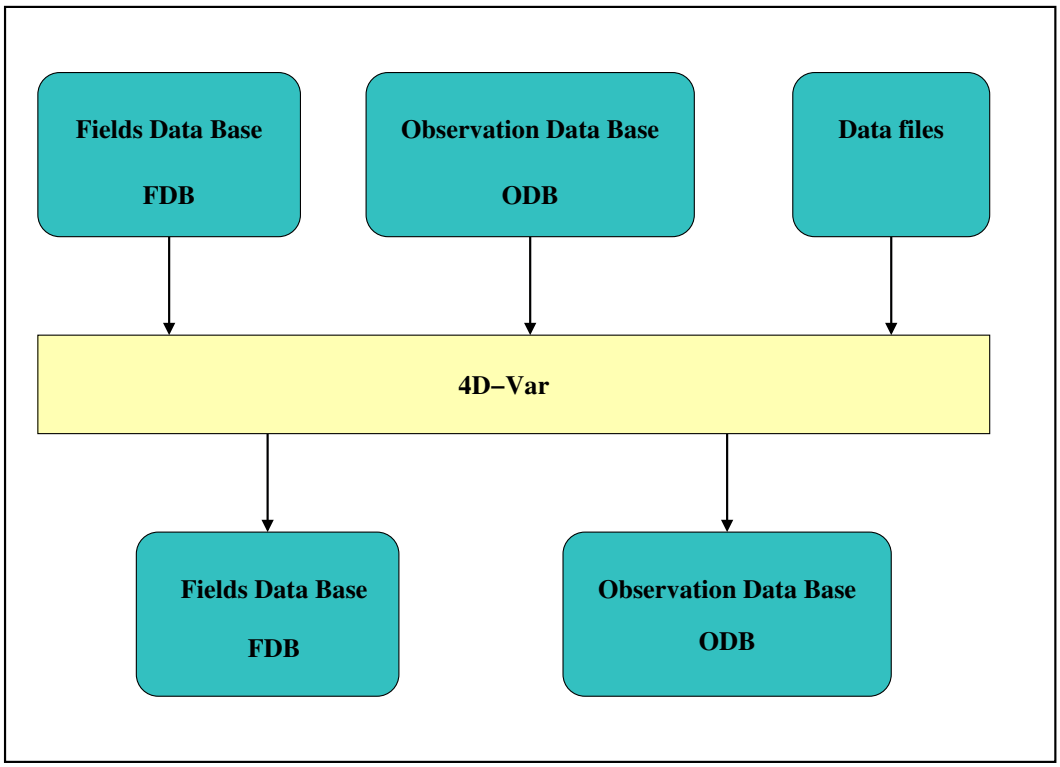

Figure 14.3 *4D-Var data flow.*

Output data from 4D-Var is in two forms.

(i) GRIB fields in the Fields Data Base, eg analysis fields, error fields.
(ii) updates to the ODB, including departures of observations from the background and the analysis and quality control information.

14.4 INPUT GRIB FIELDS

The following files in the work directory, $WDIR, contain GRIB format fields which have been extracted from the Fields Data Base. Fields of type 'fc' are taken from the forecast from the previous cycle's analysis, and are valid at the start of the 4D-Var window. Fields of type 'an' are taken from the previous analysis. The spectral orography is taken from the climatology file of the appropriate resolution, to ensure that the orography is not changed by encoding into and decoding from GRIB. Fields of type 'ef', forecast error, were output from the previous cycle's 4D-Var analysis.

Files which are needed as input to the coupled wave model also have the stream defined. If stream = DA, then the fields were output from the atmospheric model, and the GRIB codes are defined in (ECMWF local table 2, Version 128). If stream = WAVE, then the fields were output from the wave model and the GRIB codes are defined in (ECMWF local table 2, Version 140).

14.4.1 reftrajshml

Table 14.2 *Background, spherical harmonics, model levels.*

Code	Name	Description	Units	Levels	Type
129	Z	Orography (geopotential)	$m^2\ s^{-2}$	1	climate
133	Q	Specific humidity	$kg\ kg^{-1}$	1-$LEVELS	fc
130	T	Temperature	K	1-$LEVELS	fc
138	VO	Vorticity	s^{-1}	1-$LEVELS	fc
152	LNSP	Logarithm of surface pressure		1	fc
155	D	Divergence	s^{-1}	1-$LEVELS	fc
203	O3	Ozone mass mixing ratio	$kg\ kg^{-1}$	1-$LEVELS	fc

14.4.2 reftrajggml

Table 14.3 *Background, Gaussian grid, model levels.*

Code	Name	Description	Units	Levels	Type
133	Q	Specific humidity	$kg\ kg^{-1}$	1-$LEVELS	fc
203	O3	Ozone mass mixing ratio	$kg\ kg^{-1}$	1-$LEVELS	fc
246	CLWC	Cloud liquid water content	$kg\ kg^{-1}$	1-$LEVELS	fc
247	CIWC	Cloud ice water content	$kg\ kg^{-1}$	1-$LEVELS	fc
248	CC	Cloud cover	(0-1)	1-$LEVELS	fc

14.4.3 reftrajsfc

Table 14.4 *Surface fields, Gaussian grid, background (variable fields) and analysis (invariant fields).*

Code	Name	Description	Units	Type
031	CI	Sea-ice cover	(0-1)	fc
032	ASN	Snow albedo	(0-1)	fc
033	RSN	Snow density	kg m^{-3}	fc
034	SST	Sea surface temperature	K	fc
035	ISTL1	Ice surface temperature, layer 1	K	fc
036	ISTL2	Ice surface temperature, layer 2	K	fc
037	ISTL3	Ice surface temperature, layer 3	K	fc
038	ISTL4	Ice surface temperature, layer 4	K	fc
039	SWVL1	Volumetric soil water, layer 1	m^3 m^{-3}	fc
040	SWVL2	Volumetric soil water, layer 2	m^3 m^{-3}	fc
041	SWVL3	Volumetric soil water, layer 3	m^3 m^{-3}	fc
042	SWVL4	Volumetric soil water, layer 4	m^3 m^{-3}	fc
139	STL1	Soil temperature level 1	K	fc
141	SD	Snow depth	m of water equivalent	fc
148	CHNK	Charnock parameter		fc
170	STL2	Soil temperature level 2	K	fc
183	STL3	Soil temperature level 3	K	fc
198	SRC	Skin reservoir content	m of water	fc
235	SKT	Skin temperature	K	fc
236	STL4	Soil temperature level 4	K	fc
238	TSN	Temperature of snow layer	K	fc
027	CVL	Low vegetation cover	(0-1)	an
028	CVH	High vegetation cover	(0-1)	an
029	TVL	Type of low vegetation	Table index	an
030	TVH	Type of high vegetation	Table index	an
074	SDFOR	Standard deviation of filtered sub-gridscale orography		an
160	SDOR	Standard deviation of orography		an
161	ISOR	Anisotrophy of sub-gridscale orography		an
162	ANOR	Angle of sub-gridscale orography	rad	an
163	SLOR	Slope of sub-gridscale orography		an
172	LSM	Land-sea mask	(0, 1)	an
173	SR	Surface roughness	m	an
174	ALB	Albedo	(0-1)	an
234	LSRH	Logarithm of surface roughness		an

14.4.4 errgrib

Table 14.5 *Background errors, model levels, Gaussian grid.*

Code	Name	Description	Units	Levels	Type
130	T	Temperature	K	1-$LEVELS	ef
131	U	u velocity	m s^{-1}	1-$LEVELS	ef
132	V	v velocity	m s^{-1}	1-$LEVELS	ef
133	Q	Specific humidity	kg kg^{-1}	1-$LEVELS	ef
138	VO	Vorticity	s^{-1}	1-$LEVELS	ef
152	LNSP	Logarithm of surface pressure		1	ef
156	GH	Geopotential height	m	1-$LEVELS	ef
157	R	Relative humidity	%	1-$LEVELS	ef
194	BTMP	Brightness temperature	K	channels 1-54	ef
203	O3	Ozone mass mixing ratio	kg kg^{-1}	1-$LEVELS	ef

14.4.5 SSTDATA${starttime}

Table 14.6 *Background, surface, Gaussian grid, input for wave model.*

Code	Name	Description	Units	Type	Stream
031	CI	Sea-ice cover	(0-1)	fc	DA

14.4.6 SGS${starttime}

Table 14.7 *Background, surface, regular latitude/longitude grid, input for wave model.*

Code	Name	Description	Units	Type	Stream
251	2DFD	2D wave spectra	m^2 s radian^{-1}	fc	WAVE

14.4.7 CD${starttime}

Table 14.8 *Background, surface, regular latitude/longitude grid, input for wave model.*

Code	Name	Description	Units	Type	Stream
233	CDWW	Coefficient of drag with waves		fc	WAVE

14.4.8 UWAVE${starttime}

Table 14.9 *Background, surface, regular latitude/longitude grid, input for wave model.*

Code	Name	Description	Units	Type	Stream
245	WIND	10 metre wind speed	m s^{-1}	fc	WAVE

14.4.9 WIND${starttime}

Table 14.10 *Background, surface, Gaussian grid, input for wave model.*

Code	Name	Description	Units	Type	Stream
165	10U	10 metre U wind component	m s^{-1}	fc	DA
166	10V	10 metre U wind component	m s^{-1}	fc	DA

14.5 INPUT OBSERVATION DATA

Observations are read into 4D-Var from the Observation Data Base (ODB). The observation processing is described in more detail in the ODB documentation (file:///home/rd/mps/public/ugodb.pdf).

14.6 INPUT DATA FILES

Input data files can be split into two categories, invariant and date-dependent. For the invariant files, a single copy is used for the lifetime of the experiment or the operational suite. The file is copied or linked into the experiment's ${DATA} directory at start-up time in task datalinks.

Some files, such as blacklists and bias files, are date-dependent. In the operational suite, the blacklist can be changed at short notice if, for example, a satellite channel fails or a new data source arrives which has to be passively monitored to assess its quality before it can be used actively. Date-dependent files are copied to the ${WDIR} directory in task vardata at the beginning of each data assimilation cycle.

14.6.1 Invariant data files

- ${DATA}/an/cmod.table.ieee – scatterometer coefficients
- ${DATA}/an/external_bl_mon_monit.b – external blacklist file
- ${DATA}/an/moderr.cov – model error covariances for weak-constraint 4D-Var
- ${DATA}/an/neuroflux_l${LEVELS} – extended linearized longwave radiation
- ${DATA}/an/radjacobian_l${LEVELS} – extended linearized longwave radiation
- ${DATA}/an/rs_bias_T_table1 – radiosonde temperature bias correction coefficients
- ${DATA}/an/rs_bias_T_table2 – radiosonde temperature bias correction coefficients
- ${DATA}/an/rs_bias_T_table3 – radiosonde temperature bias correction coefficients
- ${DATA}/an/rszcoef_fmt – radiosonde height observation error correlation coefficients
- ${DATA}/an/ship_anemometer_heights – ship anemometer heights
- ${DATA}/an/stabal96.bal – background error balance parameters
- ${DATA}/an/stabal96.cv – background error correlations
- ${DATA}/an/wavelet_T${RESOLINC_n}_L${LEVELS}.cv – wavelet Jb background error covariances
- ${DATA}/climate/${RESOL}${GTYPE}/O3CHEM${MM} – monthly ozone chemistry climate files
- ${DATA}/ifs/namelist_${IFS_CYCLE} – an empty copy of all the IFS namelists
- ${DATA}/ifs/rtable${GTYPE}${RESOL} – namelist NAMRGRI, defining the number of points on each row of the Gaussian grid
- ${DATA}/ifs/vtable_L${LEVELS} – namelist NAMVV1, defining the hybrid vertical coordinate level coefficients
- ${DATA}/sat/amv_bias_info – atmospheric motion vector information
- ${DATA}/sat/biascoef.ssmi – SSMI bias coefficients
- ${DATA}/sat/bcor_reo3 – ozone bias correction file
- ${DATA}/sat/chanspec_noaa – noaa channel specification file
- ${DATA}/sat/cor_t_norm${LEVELS} – vertical correlation matrices for temperature for 1D-Var rain calculations
- ${DATA}/sat/cor_q_norm${LEVELS} – vertical correlation matrices for specific humidity for 1D-Var rain calculations

- ${DATA}/sat/cor_r_norm${LEVELS} – vertical correlation matrices for relative humidity for 1D-Var rain calculations
- ${DATA}/sat/cstlim_noaa – noaa cost limit file
- ${DATA}/sat/filbiaso_[ssmi|tmi] – SSMI/TMI 1D-Var bias file
- ${DATA}/sat/filcmix_[ssmi|tmi] – SSMI/TMI 1D-Var coefficient file
- ${DATA}/sat/filcovb – SSMI 1D-Var coefficient file
- ${DATA}/sat/filcovo_[ssmi|tmi] – SSMI/TMI 1D-Var coefficient file
- ${DATA}/sat/filcwat_[ssmi|tmi] – SSMI/TMI 1D-Var coefficient file
- ${DATA}/sat/mask_asc – radiosonde mask
- ${DATA}/sat/mietable_dmsp_ssmi – optical properties of hydrometeors used in the scattering calculations for 1D-Var rain
- ${DATA}/sat/rmtberr_[noaa|airs] – measurement error files
- ${DATA}/sat/rttov8/rtcoef_${platform}_${instrument}.dat – RTTOV8 radiative transfer coefficient files, for all current and historic satellite platforms and instruments
- ${DATA}/sat/scanbias.ssmi – SSMI scan bias coefficients
- ${DATA}/sat/[sigmab|correl] – background error files
- ${DATA}/sat/ssmi_tovs1c_buf – ATOVS BUFR template for conversion of SSMI data to 1c-radiances
- ${DATA}/sat/thin_reo3 – ozone thinning file
- ${DATA}/scat/mle_norm.dat – QuikSCAT look-up tables
- ${DATA}/scat/nscat2.noise – QuikSCAT noise look-up tables
- ${DATA}/scat/nscat2.table – QuikSCAT GMF look-up tables
- ${DATA}/scat/qscat1.table – QuikSCAT GMF look-up tables

14.6.2 Date-dependent data files

- ${WDIR}/bl_data_sel – data selection blacklist
- ${WDIR}/monthly_bl_mon_monit.b – monthly monitoring blacklist
- ${WDIR}/vardir/bcor_airs.dat – AIRS bias correction
- ${WDIR}/vardir/bcor_fy2.dat – FY2 bias correction
- ${WDIR}/vardir/bcor_goes.dat – GOES bias correction
- ${WDIR}/vardir/bcor_meto.dat – METEOSAT bias correction
- ${WDIR}/vardir/bcor_metop.dat – METOP bias correction
- ${WDIR}/vardir/bcor_mtst.dat – MTSAT bias correction
- ${WDIR}/vardir/bcor_noaa.dat – ATOVS bias correction
- ${WDIR}/vardir/bcor_ssmi.dat – SSM/I bias correction

14.7 OUTPUT GRIB FIELDS

Fields of type '4v' (4D-Var analysis), 'an' (analysis), 'ea' (analysis errors) and 'ef' (forecast errors) are written in GRIB code to the Fields Data Base from 4D-Var. Fig. 14.4 illustrates the difference between type '4v' and type 'an' analysis fields. For type '4v' fields, the analysis increment from the final minimization is interpolated back to high resolution and added to the penultimate high resolution trajectory at its starting point. Analysis fields output from the final high resolution non-linear trajectory are of type '4v', with a base time at the start of the trajectory and a step corresponding to the number of hours into the trajectory. So, for example, for the 1200 UTC 12-hour 4D-Var for date yyyymmdd, with an observation window from 0300 to 1500 UTC, the 4D-Var analysis at 1200 UTC is stored in the Fields Data Base and MARS with parameters:

$$\text{date} = \text{yyyymmdd}, \text{hour} = 03, \text{step} = 9, \text{type} = 4\text{v}$$

For type 'an' fields, the increment from the final minimization is added to the penultimate high resolution trajectory at the actual analysis time. Fields from the surface analysis are combined with fields from 4D-Var to give the full analysis. For the 1200 UTC 12-hour 4D-Var for date yyyymmdd, with an observation window from 0300 to 1500 UTC, the type 'an' analysis at 1200 UTC is stored in the Fields Data Base

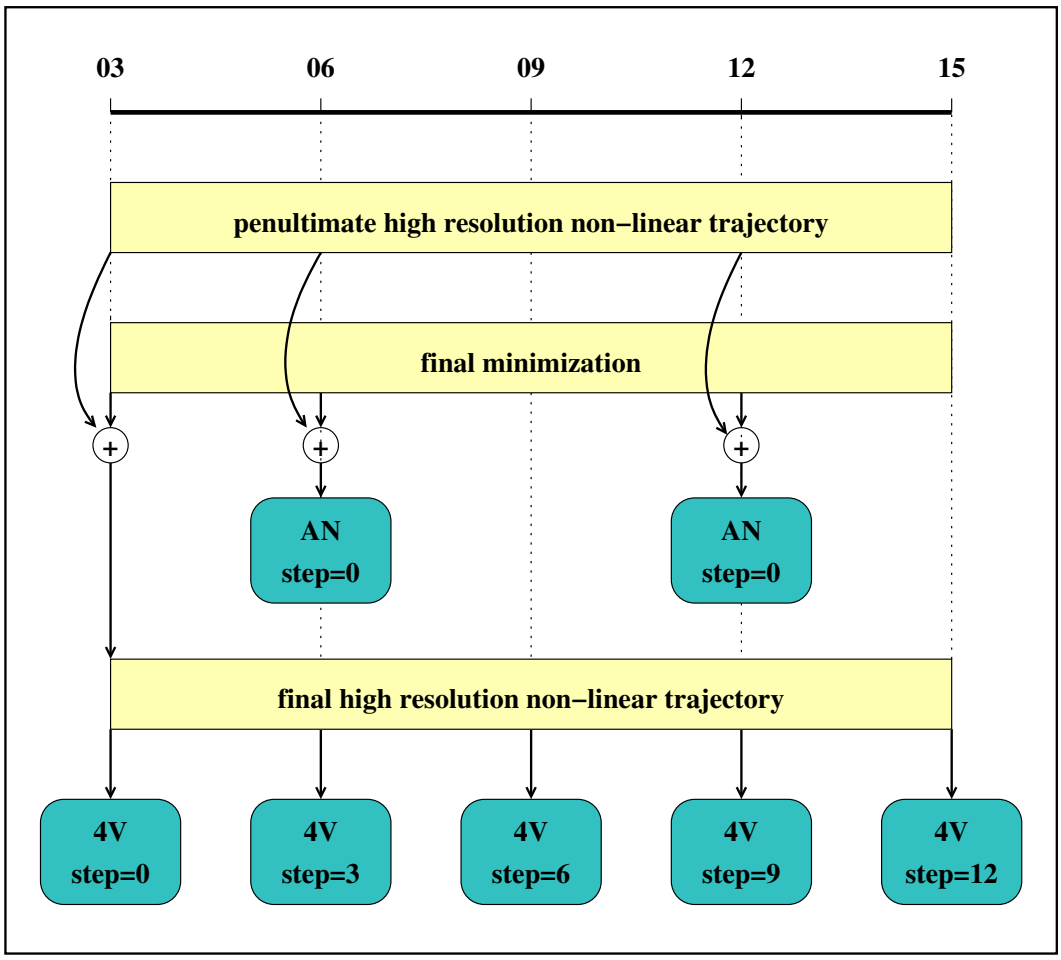

Figure 14.4 *Type 'an' and type '4v' fields written from 4D-Var to the Fields Data Base.*

and MARS with parameters:

$$\text{date} = \text{yyyymmdd}, \text{hour} = 12, \text{step} = 0, \text{type} = \text{an}$$

Output analysis fields, of type '4v' and 'an', can be generated on model levels, pressure levels and isentropic surfaces. Namelist NAMFPC controls the content of the post-processing, and there is a wide selection of fields that can be produced. The IFS determines internally whether fields should be generated in spectral or grid-point form. Described below is only the list of fields that are needed as input for the next forecast in order to cycle the data assimilation forward in time. The forecast starts from fields of type 'an'. Some of the surface fields are generated by the surface analysis jobs, which run at the same time as 4D-Var. It is important that these fields are excluded from the type 'an' post-processing of 4D-Var, so that they cannot overwrite the surface analysis fields.

14.7.1 Output type 'an' model level spectral fields

Table 14.11 *Output type 'an' model level spectral fields.*

Code	Name	Description	Units	Levels
130	T	Temperature	K	1-$LEVELS
138	VO	Vorticity	s^{-1}	1-$LEVELS
152	LNSP	Logarithm of surface pressure		1
155	D	Divergence	s^{-1}	1-$LEVELS

14.7.2 Output type 'an' model level Gaussian grid-point fields

Table 14.12 *Output type 'an' model level Gaussian grid-point fields.*

Code	Name	Description	Units	Levels
133	Q	Specific humidity	kg kg^{-1}	1-$LEVELS
203	O3	Ozone mass mixing ratio	kg kg^{-1}	1-$LEVELS
246	CLWC	Cloud liquid water content	kg kg^{-1}	1-$LEVELS
247	CIWC	Cloud ice water content	kg kg^{-1}	1-$LEVELS
248	CC	Cloud cover	(0-1)	1-$LEVELS

14.7.3 Output type 'an' surface Gaussian grid-point fields

Table 14.13 *Output type 'an' surface Gaussian grid-point fields.*

Code	Name	Description	Units
027	CVL	Low vegetation cover	(0-1)
028	CVH	High vegetation cover	(0-1)
029	TVL	Type of low vegetation	Table index
030	TVH	Type of high vegetation	Table index
032	ASN	Snow albedo	(0-1)
035	ISTL1	Ice surface temperature, layer 1	K
036	ISTL2	Ice surface temperature, layer 2	K
037	ISTL3	Ice surface temperature, layer 3	K
038	ISTL4	Ice surface temperature, layer 4	K
042	SWVL4	Volumetric soil water, layer 4	m^3 m^{-3}
074	SDFOR	Standard deviation of filtered subgrid orography	
129	Z	Orography (geopotential)	m^2 s^{-2}
148	CHNK	Charnock parameter	
160	SDOR	Standard deviation of orography	
161	ISOR	Anisotrophy of sub-gridscale orography	
162	ANOR	Angle of sub-gridscale orography	rad
163	SLOR	Slope of sub-gridscale orography	
165	10U	10 metre U wind component	m s^{-1}
166	10V	10 metre V wind component	m s^{-1}
172	LSM	Land-sea mask	(0,1)
173	SR	Surface roughness	m
174	ALB	Albedo	(0-1)
198	SRC	Skin reservoir content	m of water
234	LSRH	Logarithm of surface roughness	
235	SKT	Skin temperature	K
236	STL4	Soil temperature level 4	K

14.7.4 Output type 'an' wave model fields

The output wave model fields are on a regular latitude/longitude grid. They are identified by stream 'WAVE' and are encoded with GRIB codes defined in ECMWF local table 2, Version 140.

Table 14.14 *Output type 'an' wave model fields.*

Code	Name	Description	Units
251	2DFD	2D wave spectra	m^2 s radian^{-1}
233	CDWW	Coefficient of drag with waves	
245	WIND	10 metre wind speed	m s^{-1}

14.7.5 Output error fields

The forecast errors output from one cycle are used as the background errors input to the next cycle, and their content is described in Section 14.4.4 above. The analysis errors contain similar fields, but are of type 'ea'. The analysis errors are used to calculate the perturbations for the Ensemble Prediction System.

14.8 OUTPUT OBSERVATION DATA

Departures of observations from the background and the analysis, and quality information are written to the Observation Data Base (ODB). The observation processing is described in more detail in Part I.

14.9 SEA SURFACE TEMPERATURE ANALYSIS

The sea surface temperature analysis is done every 6 hours.

14.9.1 Input GRIB fields on model grid

These are extracted from the Fields Data Base. The background fields, of type 'fc', are taken from the forecast from the previous 4D-Var analysis time. The persistence analysis, of type 'an', is taken from the previous sea surface temperature analysis time, 6 hours earlier (which is not necessarily a 4D-Var analysis time). In the table below, 'T' is used to denote the analysis time, and 'T-6' is used to denote the persistence analysis time.

Table 14.15 *Input GRIB fields on model grid.*

Code	Name	Description	Units	Type	Time
031	CI	Sea-ice cover	(0-1)	fc	T
034	SST	Sea surface temperature	K	fc	T
034	SST	Sea surface temperature	K	an	T-6

14.9.2 Input NCEP fields

The sea surface temperature analysis and sea ice analysis from NCEP (National Center for Environmental Prediction, Washington) are input on a 0.5×0.5 degree regular latitude/longitude grid. These fields are extracted from MARS with parameters:

$$\text{expver} = 1, \text{origin} = \text{kwbc}$$

14.9.3 Input data files

- $DATA/sst/lsfil – land/sea mask for NCEP data, 0.5×0.5 degree regular latitude/longitude grid, 720×360 points
- $DATA/sst/sst_clim – surface air temperature monthly climatology, reduced to mean-sea level (0.5×0.5 degree regular latitude/longitude grid, 720×361 points)
- ${DATA}/climate/${RESOL}${GTYPE}/lsmoro – land/sea mask and orography on model Gaussian grid
- ${DATA}/climate/${RESOL}${GTYPE}/clake – lake mask

14.9.4 Output GRIB fields on model grid

The following fields are written to the Fields Data Base:

Table 14.16 *Output GRIB fields on model grid.*

Code	Name	Description	Units	Type
031	CI	Sea-ice cover	(0-1)	an
034	SST	Sea surface temperature	K	an

14.10 2 METRE TEMPERATURE ANALYSIS

The 2 metre temperature analysis is done every 6 hours.

14.10.1 Input GRIB fields on model grid

These are extracted from the Fields Data Base. The background fields, of type 'fc', are taken from the forecast from the previous 4D-Var analysis time. The invariant fields, of type 'an', are taken from the previous 4D-Var analysis.

Table 14.17 *Input GRIB fields on model grid.*

Code	Name	Description	Units	Type
129	Z	Orography	$m^2\ s^{-2}$	an
172	LSM	Land/sea mask	(0-1)	an
139	STL1	Soil temperature level 1	K	fc
167	2T	2 metre temperature	K	fc
168	2D	2 metre dewpoint temperature	K	fc

14.10.2 Input observation data

Observations are read from the Observation Data Base.

14.10.3 Output GRIB field on model grid

The analysed 2 metre temperature field is written to the Fields Data Base.

Table 14.18 *Output GRIB field on model grid.*

Code	Name	Description	Units	Type
167	2T	2 metre temperature	K	an

14.11 2 METRE RELATIVE HUMIDITY ANALYSIS

The 2 metre relative humidity analysis is done every 6 hours. Although the analysed field is 2 metre relative humidity, the final output product is 2 metre dewpoint temperature. The 2 metre relative humidity analysis cannot start until the 2 metre temperature analysis has completed, since the output from the 2 metre temperature analysis is needed in the computation of the 2 metre dewpoint temperature.

14.11.1 Input GRIB fields on model grid

These are extracted from the Fields Data Base. The background fields, of type 'fc', are taken from the forecast from the previous 4D-Var analysis time. The invariant fields, of type 'an', are taken from the previous 4D-Var analysis.

Table 14.19 *Input GRIB fields on model grid.*

Code	Name	Description	Units	Type
129	Z	Orography	$m^2\ s^{-2}$	an
172	LSM	Land/sea mask	(0-1)	an
139	STL1	Soil temperature level 1	K	fc
167	2T	2 metre temperature	K	fc
168	2D	2 metre dewpoint temperature	K	fc

14.11.2 Input observation data

Observations are read from the Observation Data Base.

14.11.3 Output GRIB field on model grid

The derived 2 metre dewpoint temperature field is written to the Fields Data Base.

Table 14.20 *Output GRIB field on model grid.*

Code	Name	Description	Units	Type
168	2D	2 metre dewpoint temperature	K	an

14.12 SNOW ANALYSIS

The snow analysis is done every 6 hours. It cannot start until the 2 metre temperature analysis has completed, since the 2 metre temperature analysis field is one of the inputs to the snow analysis.

14.12.1 Input GRIB fields on model grid

These are extracted from the Fields Data Base. The background fields, of type 'fc', are taken from the forecast from the previous 4D-Var analysis time. The invariant fields of type 'an', orography and land/sea mask, are taken from the previous 4D-Var analysis. This is denoted T4V in the table below. The persistence snow depth analysis, of type 'an', is taken from the previous snow analysis time, 6 hours earlier (which is not necessarily a 4D-Var analysis time). In the table below, 'T' is used to denote the snow analysis time, and 'T-6' is used to denote the persistence snow analysis time.

Table 14.21 *Input GRIB fields on model grid.*

Code	Name	Description	Units	Type	Time
129	Z	Orography	$m^2\ s^{-2}$	an	T4V
172	LSM	Land/sea mask	(0-1)	an	T4V
033	RSN	Snow density	$kg\ m^{-3}$	fc	T
141	SD	Snow depth	m of water equivalent	fc	T
141	SD	Snow depth	m of water equivalent	an	T-6
167	2T	2 metre temperature	K	an	T

14.12.2 Input data files

- ${DATA}/climate/${RESOL}${GTYPE}/snow – snow depth climatology (m of water equivalent) on model Gaussian grid
- ${DATA}/climate/${RESOL}${GTYPE}/cicecap – on model Gaussian grid
- $WDIR/imssnow – NESDIS snow cover field (0,1) on polar stereographic grid of approximately 25 km resolution. The data is in BUFR format, with triplets of latitude/longitude/snow cover. The NESDIS snow cover field is only used once per day, for the 06Z snow analysis.

14.12.3 Input observation data

Observations are read from the Observation Data Base.

14.12.4 Output GRIB fields on model grid

The following fields are written to the Fields Data Base:

Table 14.22 *Output GRIB fields on model grid.*

Code	Name	Description	Units	Type
033	RSN	Snow density	kg m^{-3}	an
141	SD	Snow depth	m of water equivalent	an

14.13 SOIL MOISTURE ANALYSIS

The soil moisture analysis is done every 6 hours. It cannot start until the sea surface temperature analysis, the snow analysis and 4D-Var have completed.

14.13.1 Input GRIB fields on model grid

These are extracted from the Fields Data Base. The background fields, of type 'fc', are taken from the forecast from the previous 4D-Var analysis time. The analysed fields, of type 'an', are output from the current 4D-Var analysis, the sea surface temperature analysis, the 2 metre temperature analysis, the 2 metre relative humidity analysis or the snow analysis.

Table 14.23 *Input GRIB fields on model grid.*

Code	Name	Description	Units	Type	Origin
039	SWVL1	Volumetric soil water, layer 1	$m^3\ m^{-3}$	fc	Forecast
040	SWVL2	Volumetric soil water, layer 2	$m^3\ m^{-3}$	fc	Forecast
041	SWVL3	Volumetric soil water, layer 3	$m^3\ m^{-3}$	fc	Forecast
133	Q	Specific humidity on lowest model level	$kg\ kg^{-1}$	fc	Forecast
139	STL1	Soil temperature, level 1	K	fc	Forecast
142	LSP	Large scale precipitation	m	fc	Forecast
143	CP	Convective precipitation	m	fc	Forecast
167	2T	2 metre temperature	K	fc	Forecast
168	2D	2 metre dewpoint temperature	K	fc	Forecast
170	STL2	Soil temperature, level 2	K	fc	Forecast
176	SSR	Surface solar radiation	$W\ m^{-2}\ s$	fc	Forecast
183	STL3	Soil temperature level 3	K	fc	Forecast
238	TSN	Temperature of snow layer	K	fc	Forecast
027	CVL	Low vegetation cover	(0-1)	an	4D-Var
028	CVH	High vegetation cover	(0-1)	an	4D-Var
029	TVL	Type of low vegetation	Table index	an	4D-Var
030	TVH	Type of high vegetation	Table index	an	4D-Var
129	Z	Orography	$m^2\ s^{-2}$	an	4D-Var
133	Q	Specific humidity on lowest model level	$kg\ kg^{-1}$	an	4D-Var
141	SD	Snow depth	m of water equivalent	an	Snow analysis
165	10U	10 metre U wind component	$m\ s^{-1}$	an	4D-Var
166	10V	10 metre V wind component	$m\ s^{-1}$	an	4D-Var
167	2T	2 metre temperature	K	an	2 metre temp. anal.
168	2D	2 metre dewpoint temperature	K	an	2 metre rel. hum. anal.
172	LSM	Land-sea mask	(0, 1)	an	4D-Var
174	AL	Albedo	(0-1)	an	4D-Var

14.13.2 Output GRIB fields on model grid

The following fields are output from the soil moisture analysis and written to the Fields Data Base. Before being written, the *STL1* (soil temperature level 1) field is manipulated as follows:

(i) land values are unchanged
(ii) over sea,

$$STL1 = SST \times (1 - CI) + ISTL1 \times CI$$

where

SST = analysed sea surface temperature
CI = analysed sea ice field, which varies between 0 (open water) and 1 (full ice cover)
$ISTL1$ = background soil temperature level 1

Table 14.24 *Output GRIB fields on model grid.*

Code	Name	Description	Units
039	SWVL1	Volumetric soil water, layer 1	$m^3\ m^{-3}$
040	SWVL2	Volumetric soil water, layer 2	$m^3\ m^{-3}$
041	SWVL3	Volumetric soil water, layer 3	$m^3\ m^{-3}$
139	STL1	Soil temperature, level 1	K
170	STL2	Soil temperature, level 2	K
183	STL3	Soil temperature, level 3	K
238	TSN	Temperature of snow layer	K

14.13.3 Invariant climatological fields

The final step of the soil moisture analysis task is to copy the invariant fields from the climatology files to the analysis, after first manipulating the GRIB headers to give values appropriate for the current data assimilation cycle. In this way, it is ensured that invariant fields remain unchanged, without any loss of precision due to repeatedly encoding and decoding GRIB fields.

Table 14.25 *Invariant climatological fields.*

Code	Name	Description	Units
027	CVL	Low vegetation cover	(0-1)
028	CVH	High vegetation cover	(0-1)
029	TVL	Type of low vegetation	Table index
030	TVH	Type of high vegetation	Table index
129	Z	Orography	$m^2\ s^{-2}$
160	SDOR	Standard deviation of orography	
161	ISOR	Anisotrophy of sub-gridscale orography	
162	ANOR	Angle of sub-gridscale orography	rad
163	SLOR	Slope of sub-gridscale orography	
172	LSM	Land-sea mask	(0,1)
173	SR	Surface roughness	m
174	ALB	Albedo	(0-1)
234	LSRH	Logarithm of surface roughness	

References

Alduchov, O. A. and Eskridge, R. E. (1996). Improved Magnus form approximation of saturation vapor pressure. *J. Appl. Meteorol.*, **35**, 601–609.

Andersson, E., Beljaars, A., Bidlot, J., Miller, M., Simmons, A. and Thépaut, J. N. (2003). A major new cycle of the IFS: Cycle 25r4. *ECMWF Newsletter*, **97**, 12–20.

Andersson, E., Cardinali, C., Isaksen, L. and Garcia-Mendez, A. (2001). On the impact of frequent data in ECMWF's 4D-Var scheme: Hourly surface pressure data, European profilers and profiling aircraft data. In *Proc. of the 8th ECMWF Workshop on Meteorological Operational Systems*, pp. 179–183.

Andersson, E., Fisher, M., Munro, R. and McNally, A. (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. *Q. J. R. Meteorol. Soc.*, **126**, 1455–1472.

Andersson, E. and Garcia-Mendez, A. (2002). Assessment of European wind profiler data, in an NWP context. *ECMWF Tech. Memo. No. 372.*

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljevic, D., Brankovic, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J.-N. and Viterbo, P. (1998). The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part III: Experimental results. *Q. J. R. Meteorol. Soc.*, **124**, 1831–1860.

Andersson, E., Hólm, E. and Thépaut, J. N. (2004). Impact studies of main types of conventional and satellite humidity data. In *Proc. 3rd WMO Workshop on The Impact of Various Observing Systems on NWP*, Alpbach, Austria, 9–12 March 2004.

Andersson, E. and Järvinen, H. (1999). Variational quality control. *Q. J. R. Meteorol. Soc.*, **125**, 697–722.

Andersson, E., Pailleux, J., Thépaut, J.-N., Eyre, J. R., McNally, A. P., Kelly, G. A. and Courtier, P. (1994). Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **120**, 627–653.

Attema, E. P. W. (1986). An experimental campaign for the determination of the radar signature of the ocean at C-band. In *Proc. Third International Colloquium on Spectral Signatures of Objects in Remote Sensing*, pp. 791–799, Les Arcs, France, ESA, SP-247.

Auligné, T. and Dee, D. (2006). Varbc user guide. *Technical report*, available from Thomas Auligné, ECMWF.

Bauer, P., Lopez, P., Benedetti, A., Salmond, D. and Moreau, E. (2006a). Implementation of 1D+4D-Var assimilation of precipitation affected microwave radiances at ECMWF. *Q. J. R. Meteorol. Soc.*, **132**, 2277–2306.

Bauer, P., Lopez, P., Benedetti, A., Salmond, D., Saarinen, S. and Bonazzola, M. (2006b). Implementation of 1D+4D-Var assimilation of precipitation affected microwave radiances at ECMWF, Part II: 4D-Var. *Q. J. R. Meteorol. Soc.*, **132**, 2307–2332.

Bauer, P., Moreau, E., Chevallier, F. and O'Keefe, U. (2006c). Multiple-scattering microwave radiative transfer for data assimilation applications. *Q. J. R. Meteorol. Soc.*, **132**, 1259–1281.

Benedetti, A. and Fisher, M. (2006). Background error statistics for aerosols. *ECMWF Tech. Memo. No. 489.*

Betts, A. (1997). The parametrization of deep convection: A review. In *Proc. ECMWF Workshop on New Insights and Approaches to Convective Parametrization*, pp. 166–188, 4–7 November 1996.

Blackadar, A. K. (1962). The vertical distribution of wind and turbulent exchange in a neutral atmosphere. *J. Geophys. Res.*, **67**, 3095–3102.

Blondin, C. (1991). Parametrization of land surface processes in numerical weather prediction. In T. J. Schmugge and J.-C. André (Eds), *Land Surface Evaporation: Measurement and Parametrization*, pp. 31–54, Springer-Verlag.

Bormann, N. and Healy, S. (2006). A fast radiative-transfer model for the assimilation of MIPAS limb radiances: Accounting for horizontal gradients. *Q. J. R. Meteorol. Soc.*, **132**, 2357–2376.

Bormann, N., Healy, S. and Hamrud, M. (2007). Assimilation of MIPAS limb radiances in the ECMWF system. Part II: Experiments with a 2-dimensional observation operator and comparison to retrieval assimilation. *Q. J. R. Meteorol. Soc.*, **133**, in press.

Bormann, N., Matricardi, M. and Healy, S. (2005). A fast radiative transfer model for the assimilation of infrared limb radiances from MIPAS. *Q. J. R. Meteorol. Soc.*, **131**, 1631–1653.

Bormann, N., Saarinen, S., Kelly, G. and Thépaut, J.-N. (2003). The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Mon. Wea. Rev.*, **131**, 706–718.

Bormann, N. and Thépaut, J.-N. (2007). Assimilation of MIPAS limb radiances in the ECMWF system. Part I: Experiments with a 1-dimensional observation operator. *Q. J. R. Meteorol. Soc.*, **133**, in press.

Bouttier, F. (2001a). The development of 12-hourly 4D-Var. *ECMWF Tech. Memo. No. 348.*

Bouttier, F. (2001b). The use of profiler data at ECMWF. *Meteorologische Zeitschrift*, **10**, 497–510.

Bouttier, F., Derber, J. and Fisher, M. (1997). The 1997 revision of the Jb term in 3D/4D-Var. *ECMWF Tech. Memo. No. 238.*

Buck, A. L. (1981). New equations for computing vapor pressure and enhancement factor. *J. Appl. Meteorol.*, **20**, 1527–1532.

Buizza, R. (1994). Sensitivity of optimal unstable structures. *Q. J. R. Meteorol. Soc.*, **120**, 429–451.

Cardinali, C., Andersson, E., Viterbo, P., Thépaut, J.-N. and Vasiljevic, D. (1994). Use of conventional surface observations in three-dimensional variational data assimilation. *ECMWF Tech. Memo. No. 205.*

Cardinali, C., Isaksen, L. and Andersson, E. (2003). Use and impact of automated aircraft data in a global 4D-Var data assimilation system. *Mon. Wea. Rev.*, **131**, 1865–1877.

Cavalieri, D. J., Crawford, J. P., Drinkwater, M. R., Eppler, D. T., Farmer, L. D., Jentz, R. R. and Wackerman, C. C. (1991). Aircraft active and passive microwave validation of sea-ice concentration from the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager. *J. G. R. Oceans*, **96**(C12), 21989–22008.

Chelton, D. B., Freilich, M. H. and Esbensen, S. K. (2000). Satellite observations of the wind jets off the Pacific coast of Central America. Part I: Case studies and statistical characteristics. *Mon. Wea. Rev.*, **128**, 1993–2018.

Chevallier, F. and Mahfouf, J.-F. (2001). Evaluation of the Jacobians of infrared radiation models for variational data assimilation. *J. Appl. Meteorol.*, **40**, 1445–1461.

Chevallier, F., Morcrette, J.-J., Chéruy, F. and Scott, N. A. (2000). Use of a neural network-based longwave radiative transfer scheme in the ECMWF atmospheric model. *Q. J. R. Meteorol. Soc.*, **126**, 761–776.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M. (1998). The ECMWF implementation of three dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.*, **124**, 1783–1807.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367–1388.

Dee, D. (2004). Variational bias correction of radiance data in the ECMWF system. In *Proc. of the ECMWF Workshop on Assimilation of High Spectral Resolution Sounders in NWP*, pp. 97–112.

Derber, J. C. and Bouttier, F. (1999). A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**, 195–221.

Dethof, A. and Holm, E. (2004). Ozone assimilation in the ERA-40 reanalysis project. *Q. J. R. Meteorol. Soc.*, **130**, 2851–2872.

Douville, H., Viterbo, P., Mahfouf, J.-F. and Beljaars, A. C. M. (2001). Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data. *Mon. Wea. Rev.*, **128**, 1733–1756.

Drusch, M. (2006). Sea ice concentration analyses for the Baltic Sea and their impact on numerical weather prediction. *J. Appl. Meteorol. and Clim.*, **45**(7), 982–994.

Drusch, M., Vasiljevic, D. and Viterbo, P. (2004). ECMWF's global snow analysis: Assessment and revision based on satellite observations. *J. Appl. Meteorol.*, **43**(9), 1282–1294.

Dubuisson, P. J., Buriez, J. and Fouquart, Y. (1996). High spectral solar radiative transfer in absorbing and scattering media: Application to the satellite simulation. *J. Quant. Spectrosc. Radiat. Transfer*, **55**, 103–126.

Ebert, E. E. and Curry, J. A. (1992). A parametrization of ice optical properties for climate models. *J. Geophys. Res.*, **97D**, 3831–3836.

Engelen, R. J., Andersson, E., Chevallier, F., Hollingsworth, A., Matricardi, M., McNally, A., Thépaut, J.-N. and Watts, P. (2004). Estimating atmospheric CO2 from advanced infrared satellite radiances within an operational 4D-Var data assimilation system: Methodology and first results. *J. Geophys. Res.*, **109**, D19309.

Eyre, J. R. (1991). A fast radiative transfer model for satellite sounding systems. *ECMWF Tech. Memo. No. 176.*

Fisher, M. (1996). The specification of background error variances in the ECMWF variational analysis system. In *Proc. ECMWF Workshop on Non-linear Aspects of Data Assimilation*, pp. 645–652, Reading, 9–11 September 1996.

Fisher, M. (1998). Minimization algorithms for variational data assimilation. In *Proc. ECMWF Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling*, pp. 364–385, Reading, U.K., 7–11 September 1998.

Fisher, M. (2003). Background error covariance modelling. In *Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, pp. 45–64, Reading, U.K., 8–12 Sept 2003.

Fisher, M. (2004). Generalized frames on the sphere, with application to the background error covariance modelling. In *Proc. ECMWF Seminar on Recent Developments in Numerical Methods for Atmosphereic and Ocean Modelling*, pp. 87–102, Reading, U.K., 6–10 September 2004.

Fisher, M. (2006). Wavelet jb - a new way to model the statistics of background errors. *ECMWF Newsletter*, **106**, 23–28.

Fisher, M. and Andersson, E. (2001). Developments in 4D-Var and Kalman Filtering. *ECMWF Tech. Memo. No. 347.*

Fisher, M. and Courtier, P. (1995). Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Tech. Memo. No. 220.*

Fouquart, Y. (1987). Radiative transfer in climate models. In M. E. Schlesinger (Ed.), *Physically Based Modelling and Simulation of Climate and Climate Changes*, pp. 223–284, Kluwer Acad. Publ.

Fouquart, Y. and Bonnel, B. (1980). Computations of solar heating of the earth's atmosphere: A new parametrization. *Beitr. Phys. Atmos.*, **53**, 35–62.

Gauthier, P. and Thépaut, J.-N. (2001). Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Meteo-France. *Mon. Wea. Rev.*, **129**, 2089–2102.

Geleyn, J.-F. (1988). Interpolation of wind, temperature and humidity values from the model levels to the height of meaurement. *Tellus*, **40**, 347–351.

Geleyn, J.-F. and Hollingsworth, A. (1997). An economical and analytical method for the interactions between scattering and line absorption of radiation. *Contrib. to Atmos. Phys.*, **52**, 1–16.

Gilbert, J. C. and Lemaréchal, C. (1989). Some numerical experiments with variable storage quasi-Newton algorithms. *Math. Prog.*, **B25**, 407–435.

Gloersen, P., Campbell, W. J., Cavalieri, D. J., Comiso, J. C., Parkinson, C. L. and Zwally, H. J. (1994). Arctic and Antarctic sea ice, 1978–1987: Satellite passive microwave observations and analysis. In *NASA SP511*, Greenbelt, MD, NASA.

Gloersen, P. and Cavalieri, D. J. (1986). Reduction of weather effects in the calculation of sea-ice concentration from microwave radiances. *J. G. R. Oceans*, **91**(C3), 3913–3919.

Gregory, D. and Miller, M. (1989). A numerical study of the parametrization of deep tropical convection. *Q. J. R. Meteorol. Soc.*, **115**, 1209–1241.

Grumbine, R. W. (1996). Automated passive microwave sea-ice concentration analysis at NCEP. *NOAA NCEP Tech. Note.*

Gustafsson, N. (1992). Use of a digital filter as a weak constraint in variational data assimilation. In *Proc. ECMWF Workshop on Variational Data Assimilation with Special Emphasis on 3-dimensional Aspects*, pp. 327–338, Reading, U.K.

Haseler, J. (2004). Early-delivery suite. *ECMWF Tech. Memo. No. 454.*

Healy, S. and Thépaut, J.-N. (2006). Assimilation experiments with CHAMP GPS radio occultation measurements. *Q. J. R. Meteorol. Soc.*, **132**, 605–623.

Hersbach, H., Isaksen, L., Leidner, S. M. and Janssen, P. A. E. M. (2004). The impact of SeaWinds scatterometer data in the ECMWF 4D-Var assimilation system (in preparation).

Hollinger, J., Lo, R. and Poe, G. (1987). *Special Sensor Microwave/Imager User's Guide.* Nav. Res. Lab., Washington D.C.

Hólm, E., Andersson, E., Beljaars, A., Lopez, P., Mahfouf, J.-F., Simmons, A. J. and Thépaut, J.-N. (2002). Assimilation and modeling of the hydrological cycle: ECMWF's status and plans. *ECMWF Tech. Memo. No. 383.*

Huddleston, J. N. and Stiles, B. W. (2000). Multidimensional histogram (MUDH) rain flag. *Product Description Version 2.1*, URL http://podaac-www.jpl.nasa.gov/quikscat/.

Ingleby, N. B. and Lorenc, A. C. (1993). Bayesian quality control using multivariate normal distributions. *Q. J. R. Meteorol. Soc.*, **119**, 1195–1225.

Isaksen, L. (1997). Impact of ERS scatterometer data in the ECMWF 4D-Var assimilation system. Preliminary studies. In *Space at the Service of our Environment*, pp. 1829–1851, Florence, Italy, May 1997, ESTEC, European Space Agency [SP-414]. Available from ESA Publications Division, ESTEC Noordwijk, The Netherlands.

Isaksen, L. and Janssen, P. A. E. M. (2004). Impact of ERS scatterometer winds in ECMWF's assimilation. *Q. J. R. Meteorol. Soc.*, **130**, 1793–1814.

Isaksen, L. and Stoffelen, A. C. M. (2000). ERS scatterometer wind data impact on ECMWF's tropical cyclone forecasts. *IEEE Trans. Geosci. Remote Sens.*, **38**(4), 1885–1892.

Janisková, M., Mahfouf, J.-F., Morcrette, J.-J. and Chevallier, F. (2000). Development of linearized radiation and cloud schemes for the assimilation of cloud properties. *ECMWF Tech. Memo. No. 301.*

Janisková, M., Mahfouf, J.-F., Morcrette, J.-J. and Chevallier, F. (2002). Linearized radiation and cloud schemes in the ECMWF model: Development and evaluation. *Q. J. R. Meteorol. Soc.*, **128**, 1505–1527.

Järvinen, H., Andersson, E. and Bouttier, F. (1999). Variational assimilation of time sequences of surface observations with serially correlated errors. *ECMWF Tech. Memo. No. 266.*

Järvinen, H., Saarinen, S. and Undén, P. (1996). *User's Guide for Blacklisting.* Available on request from ECMWF, Shinfield Park, RG2 9AX, Reading, Berkshire, UK.

Järvinen, H. and Undén, P. (1997). Observation screening and first guess quality control in the ECMWF 3D-Var data assimilation system. *ECMWF Tech. Memo. No. 236.*

Kelly, G., Andersson, E., Hollingsworth, A., Lönnberg, P., Pailleux, J. and Zhang, Z. (1991). Quality control of operational physical retrievals of satellite sounding data. *Mon. Wea. Rev.*, **119**, 1866–1880.

Kelly, G. and Pailleux, J. (1988). Use of satellite vertical sounder data in the ECMWF analysis system. *ECMWF Tech. Memo. No. 143.*

Klinker, E., Rabier, F., Kelly, G. and Mahfouf, J. F. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteorol. Soc.*, **126**, 1191–1215.

Köpken, C., Kelly, G. and Thépaut, J.-N. (2004). Assimilation of Meteosat radiance data within the 4D-Var system at ECMWF: Assimilation experiments and forecast impact. *Q. J. R. Meteorol. Soc.*, **130**, 2277–2292.

Leidner, S. M., Hoffman, R. N. and Augenbaum, J. (2000). *SeaWinds Scatterometer Real-Time BUFR Geophysical Data Product, User's Guide Version 2.3.0.* NOAA/NESDIS.

Leidner, S. M., Hofman, R. N. and Augenbaum, J. (1999). *SeaWinds Satterometer Real-Time BUFR Geophysical Data Product User's Guide Version 1.0.* Available from ECMWF and AER.

Leidner, S. M., Isaksen, L. and Hoffman, R. N. (2003). Impact of NSCAT winds on tropical cyclones in the ECMWF 4D-Var assimilation system. *Mon. Wea. Rev.*, **131**, 3–26.

Lönnberg, P. (1989). Developments in the ECMWF analysis system. In *Proc. ECMWF Seminar on Data assimilation and the Use of Satellite Data*, pp. 75–119, 5–9 September 1988.

Lönnberg, P. and Shaw, D. (1985). Data selection and quality control in the ECMWF analysis system. In *ECMWF Workshop on The Use And Quality Control of Meteorological Observations*, pp. 225–254, 6–9 November 1984.

Lönnberg, P. and Shaw, D. (Eds) (1987). *ECMWF Data Assimilation Scientific Documentation, Research Manual 1.*

Lopez, P. and Moreau, E. (2005). A convection scheme for data assimilation: Description and initial tests. *Q. J. R. Meteorol. Soc.*, **131**, 409–436.

Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194.

Lorenc, A. C. (1988). Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc.*, **114**, 205–240.

Lott, F. and Miller, M. J. (1997). A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Q. J. R. Meteorol. Soc.*, **123**, 101–127.

Louis, J.-F. (1979). A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteorol.*, **17**, 187–202.

Louis, J.-F., Tiedtke, M. and Geleyn, J.-F. (1982). A short history of the PBL parametrization at ECMWF. In *Proc. ECMWF Workshop on Planetary Boundary Layer Parameterization*, pp. 59–80, Reading, 25–27 November, 1981.

Lynch, P. (1993). Digital Filters for Numerical Weather Prediction. *HIRLAM Technical Report No. 10.*

Lynch, P. (1997). The Dolph-Chebyshev Window: A simple optimal filter. *Mon. Wea. Rev.*, **125**, 655–660.

Machenhauer, B. (1977). On the dynamics of gravity oscillations in a shallow water model, with application to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253–271.

Mahfouf, J.-F. (1991). Analysis of soil moisture from near surface parameters: A feasibility study. *J. Appl. Meteorol.*, **30**, 1534–1547.

Mahfouf, J.-F. (1999). Influence of physical processes on the tangent-linear approximation. *Tellus*, **51A**, 147–166.

Mahfouf, J.-F., Buizza, R. and Errico, R. M. (1997). Strategy for including physical processes in the ECMWF data assimilation system. In *Proc. ECMWF Workshop on Non-linear Aspects of Data Assimilation*, Shinfield Park, Reading, RG2 9AX, 9–11 September 1996.

Mahfouf, J. F. and Rabier, F. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with improved physics. *Q. J. R. Meteorol. Soc.*, **126**, 1171–1190.

Matricardi, M., Chevallier, F. and Tjemkes, S. (2001). An improved fast radiative transfer model for the assimilation of radiance observations. *ECMWF Tech. Memo. No. 345.*

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

Mlawer, E., Taubman, S. J., Brown, P. D., Ianoco, M. and Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663–16682.

Morcrette, J.-J. (1989). Description of the radiation scheme in the ECMWF operational weather forecast model. *ECMWF Tech. Memo. No. 165.*

Morcrette, J.-J. (1991). Radiation and cloud radiative properties in the ECMWF operational forecast model. *J. Geophys. Res.*, **96D**, 9121–9132.

Morcrette, J.-J. (1998). Description of the radiation scheme in the ECMWF operational weather forecast model. *ECMWF Tech. Memo. No. 252.*

Morcrette, J.-J. (2000). On the effects of the temporal and spatial sampling of radiation fields on the ECMWF forecasts and analyses. *Mon .Wea. Rev.*, **128**, 876–887.

Munro, R., Köpken, C., Kelly, G., Thépaut, J.-N. and Saunders, R. (2004). Assimilation of Meteosat radiance data within the 4D-Var system at ECMWF: Data quality monitoring, bias correction and single-cycle experiments. *Q. J. R. Meteorol. Soc.*, **130**, 2293–2313.

Pailleux, J. (1990). A global variational assimilation scheme and its application for using TOVS radiances. In *Proc. WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography*, pp. 325–328, Clermont-Ferrand, France.

Portabella, M. and Stoffelen, A. (2000). Quality control and wind retrieval for SeaWinds. *Scientific report WR-2002-01*, Koninklijk Nederlands Meteorologisch Instituut, The Netherlands.

Rabier, F. and Courtier, P. (1992). Four-dimensional assimilation in the presence of baroclinic instability. *Q. J. R. Meteorol. Soc.*, **118**, 649–672.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126**, 1143–1170.

Rabier, F., Mahfouf, J.-F., Fisher, M., Järvinen, H., Simmons, A., Andersson, E., Bouttier, F., Courtier, P., Hamrud, M., Haseler, J., Hollingsworth, A., Isaksen, L., Klinker, E., Saarinen, S., Temperton, C., Thépaut, J.-N., Undén, P. and Vasiljevic, D. (1997). Recent experimentation on 4D-Var and first results from a simplified Kalman filter. *ECMWF Tech. Memo. No. 240.*

Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A. and Bouttier, F. (1998). The ECMWF implementation of three dimensional variational assimilation (3D-Var). II: Structure funtions. *Q. J. R. Meteorol. Soc.*, **124**, 1809–1829.

Radnóti, G., Trémolet, Y., Andersson, E., Isaksen, L., Hólm, E. and Janisková, M. (2005). Diagnostics of linear and incremental approximations in 4d-var revisited for higher resolution. *ECMWF Tech. Memo. No. 467.*

Saunders, R. W. and Matricardi, M. (1998). A fast forward model for ATOVS (RTATOV). In *Tech. Proc. 9th International TOVS Study Conference*, p. 11 pp, Igls, Austria, 20–26 February, 1997.

Savijärvi, H. (1995). Error growth in a large numerical forecast system. *Mon. Wea. Rev.*, **123**, 212–221.

Sherlock, V. J. (1999). ISEM-6: Infrared Surface Emissivity Model for RRTOV-6. *Forecasting Research Technical Report No. 299*, available from Met Office, London Road, Bracknell, Berkshire RG12 2SZ, UK.

Simmons, A. J. and Burridge, D. (1981). An energy and angular momentum conserving vertical finite difference scheme and hybrid coordinate. *Mon. Wea. Rev.*, **109**, 758–766.

Simmons, A. J. and Chen, J. (1991). The calculation of geopotential and the pressure gradient in the ECMWF atmospheric model: Influence on the simulation of the polar atmosphere and on temperature analyses. *Q. J. R. Meteorol. Soc.*, **117**, 29–58.

Slingo, J. M. (1997). The development and verification of the cloud prediction scheme in the ECMWF model. *Q. J. R. Meteorol. Soc.*, **113**, 899–927.

Smith, E. A. and Shi, L. (1992). Surface forcing of the infrared cooling profile over the Tibetan plateau. Part I: Influence of relative longwave radiative heating at high altitude. *J. Atmos. Sci.*, **49**, 805–822.

Stoffelen, A. (1999). *Scatterometry.* Ph.D. thesis, KNMI.

Stoffelen, A. and Anderson, D. (1997). Ambiguity removal and assimilation of scatterometer data. *Q. J. R. Meteorol. Soc.*, **123**, 491–518.

Talagrand, O. (1997). Assimilation of observations, an Introduction. *J. Meteorol. Soc. Japan*, **75**, N.1B,191–209.

Temperton, C. (1988). Implicit normal mode initialization. *Mon. Wea. Rev.*, **116**, 1013–1031.

Temperton, C. (1989). Implicit normal mode initialization for spectral models. *Mon. Weather Rev.*, **117**, 436–451.

Thépaut, J.-N. and Andersson, E. (2003). Assimilation of high-resolution satellite data. *ECMWF Newsletter*, **97**, 6–12.

Tiedtke, M. (1989). A comprehensive massflux scheme for cumulus parametrization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800.

Tiedtke, M. (1993). Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121**, 3040–3061.

Tiedtke, M. (1996). An extension of cloud-radiation parametrization in the ECMWF model: The representation of subgrid-scale variations of optical depth. *Mon. Wea. Rev.*, **124**, 745–750.

Tomassini, M., LeMeur, D. and Saunders, R. (1997). Satellite wind observations of hurricanes and their impact on NWP model analyses and forecasts. *Mon. Wea. Rev. (to appear).*

Tompkins, A. and Janisková, M. (2004). A cloud scheme for data assimilation: Description and initial tests. *Q. J. R. Meteorol. Soc.*, pp. 2495–2518.

Trémolet, Y. (2003). Model error in variational data assimilation. In *Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, pp. 361–367, Reading, U.K., 8–12 Sept 2003.

Trémolet, Y. (2004). Diagnostics of linear and incremental approximations in 4D-Var. *Q. J. R. Meteorol. Soc.*, **130**, 2233–2251.

Trémolet, Y. (2005). Incremental 4d-var convergence study. *ECMWF Tech. Memo. No. 469.*

Uppala, S. M. *et al.* (2005). The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, **131**, 2961–3012.

Vasiljevic, D., Cardinali, C. and Undén, P. (1992). ECMWF 3D-Variational assimilation of conventional observations. In *Proc. ECMWF Workshop on Variational Assimilation with Emphasis on Three-dimensional Aspects*, Reading, 9–12 November 1992.

Veerse, F. and Thépaut, J. N. (1998). Multiple-truncation incremental approach for four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **124**, 1889–1908.

Wergen, W. (1987). Diabatic nonlinear normal mode initialisation for a spectral model with a hybrid vertical coordinate. *ECMWF Tech. Report No. 59.*