

# IFS DOCUMENTATION – Cy41r2

## Operational implementation 8 March 2016

### PART I: OBSERVATIONS

#### Table of contents

- [Chapter 1 Observation operators](#)
- [Chapter 2 Observation processing](#)
- [Chapter 3 Observation screening](#)
- [References](#)

© Copyright 2016

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

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF. The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.



## Chapter 1

# Observation operators

### Table of contents

#### 1.1 Introduction

- 1.1.1 Data selection controls, and the interface to the blacklist
- 1.1.2 Variable numbers and association with observation operators
- 1.1.3 Organization in observation sets
- 1.1.4 Generic parts of the observation operator

#### 1.2 Observation operators – general

#### 1.3 The observation operator for conventional data

- 1.3.1 Geopotential height
- 1.3.2 Wind
- 1.3.3 Humidity
- 1.3.4 Temperature

#### 1.4 Satellite radiance operators

- 1.4.1 Common aspects for the setup of nadir radiance assimilation
- 1.4.2 Clear-sky nadir radiances and overcast infrared nadir radiances
- 1.4.3 All-sky nadir radiances
- 1.4.4 Cloud affected infrared radiances
- 1.4.5 Clear-sky limb radiances

#### 1.5 Other satellite observation operators

- 1.5.1 Atmospheric Motion Vectors
- 1.5.2 Thicknesses
- 1.5.3 Gas retrievals
- 1.5.4 Scatterometer winds
- 1.5.5 GPS Radio Occultation bending angles

#### 1.6 Surface observation operators

- 1.6.1 Vertical interpolation
- 1.6.2 Surface values of dry static energy
- 1.6.3 Transfer coefficients
- 1.6.4 Two-metre relative humidity

### 1.1 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  $H$  signifies the ensemble of operators transforming the control variable  $\mathbf{x}$  into the equivalent of each observed quantity,  $\mathbf{y}^o$ , at observation locations. In this chapter we define the content of each of the IFS observation operators. A number of observation-related parts of the data assimilation algorithm are described in Part 2, particularly: calculation of departures and the  $J_o$  costfunction; variational quality control (VarQC); variational bias correction (VarBC); modelling correlated observation error; horizontal interpolation from model to observation locations.

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 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 3](#).

### 1.1.1 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 1.1](#)) and body ([Table 1.2](#)) 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.

### 1.1.2 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**.

Some of the traditional operator names are: 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.

**Table 1.1** Header variables in the ifs/blacklist interface. Exact contents may change - see [black.F90](#).

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	QL1	quality indicator 1
13	QL2	quality indicator 2
14	QL3	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 2-metre 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
33	SPECIFIC	amsua specific
34	SEA_ICE	model sea ice fraction

**Table 1.2** Body variables for the ifs/blacklist interface. Exact contents may change - see [black.F90](#)

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

**Table 1.3** Association between variable numbering and observation operator routines for some of the longstanding observation types. The CVAR-NAMEs also appear in the printed  $J_o$ -table in the log-file. For the full and up-to-date set of variable numbers see [yomcosjo.F90](#).

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 (M-France)
11	T2	PPT2M	2-metre temperature
12	TS		Surface temperature (M-France)
13	RAD	RADTR/RADTR_ML	Radiance data
14	SN		Snow (M-France)
15	RR		Rain rate (M-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 <sub>0</sub>
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	PP_UV(+conversion to HLOS in HOP)	Horizontal line-of-sight winds
30	AOD	AOD_OP	Aerosol optical depth
31	LRA	RTL_HOP_1D	Limb sound radiance

### 1.1.3 Organization in observation sets

The vertical observation operators are vectorized over NMXLLEN (yomdim) data. To achieve this the data first have to be sorted by type and subdivided into sets of lengths not exceeding that number. NMXLLEN 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.

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/S) must not contain data from more than one instrument. That is to say that inside [HOP](#), [HOPTL](#) and [HOPAD](#) (and routines called from 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.

### 1.1.4 Generic parts of the observation operator

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/HOPTH/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 known from the information in tables such as **MTYPOB(KSET)** stored in **yomobset**.

The following describes **HOP//HOPTH/**. 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**. Height 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 1.3.1](#). 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.
- (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** or **RADTR\_ML** which calls the radiative transfer code. Computation of FG departures is done in **HDEPART**, which is described (along with the observation costfunction) in Part 2.

## 1.2 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 converts from control variables to model variables.
- (ii) The inverse spectral transforms put the model variables on the model's reduced Gaussian grid.
- (iii) A horizontal interpolation gives vertical profiles of model variables at observation points
- (iv) Observation operators are applied, including vertical integration of, for example, the hydrostatic equation to form geopotential followed by vertical interpolation to the level of the observations (typical for conventional observations) and of the radiative transfer equation to form radiances (if applicable).

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 1.3](#)). The current geopotential vertical interpolation together with the temperature vertical interpolation are not exactly consistent with hydrostatics. A new consistent and accurate vertical interpolation has been devised by Météo-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 10-metre wind (`PPUV10M`) and 2-metre temperature (`PPT2M`) match an earlier version of the model's surface layer parametrization. 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-metre 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-metre wind, on the other hand, is used for all surface winds (SYNOP, DRIBU, TEMP, PILOT and SCAT), where interpolation is not confined to only 10 m, but is performed to the actual observation height (in practise ranging from 4 to 10 m).

Relative humidity is assumed constant in the lowest model layer to evaluate its 2-metre value (`PPRH2M`), see [Subsection 1.6.4](#). Model equivalents of total column water vapour data are obtained by vertical integration of  $q$  (in `GPPWC` and `PPPWC`). The routine `PPPWC` is also used for vertical integration of GEMS/MACC trace gasses. Observation operators exist for precipitable water content (also using `PPPWC`) and thicknesses (`PPGEOPTL`).

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.

## 1.3 THE OBSERVATION OPERATOR FOR CONVENTIONAL DATA

### 1.3.1 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

$$T_{\text{ICAO}} = T_0 - \frac{\Lambda}{g} \phi_{\text{ICAO}} \quad (1.1)$$

where  $T_0$  is 288 K,  $\phi_{\text{ICAO}}$  is the geopotential above 1013.25 hPa and  $\Lambda$  is 0.0065 K m<sup>-1</sup> 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  $T_{\text{ICAO}}$  and the geopotential  $\phi_{\text{ICAO}}$  as a function of pressure (`PPSTA`). We may then evaluate the geopotential  $\phi(p)$  at any pressure  $p$  following

$$\phi(p) - \phi_{\text{surf}} = \phi_{\text{ICAO}}(p) - \phi_{\text{ICAO}}(p_{\text{surf}}) + \Delta\phi \quad (1.2)$$

where  $p_{\text{surf}}$  is the model surface pressure and  $\phi_{\text{surf}}$ , 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

$$\Delta\phi_k = \sum_{j=L}^{k+1} R_{\text{dry}}(T_{v_j} - T_{\text{ICAO}_j}) \ln\left(\frac{p_{j+1/2}}{p_{j-1/2}}\right) + \alpha_k R_{\text{dry}}(T_{v_k} - T_{\text{ICAO}_k}) \quad (1.3)$$

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

$$z(\ln p) = a + b(\ln p) + c(\ln p)^2 \quad (1.4)$$

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

$$\phi(\ln p) = z_2 + \frac{(z_2 - z_1)(\ln p - \ln p_2)(\ln p - \ln p_3)}{(\ln p_2 - \ln p_1)(\ln p_1 - \ln p_3)} - \frac{(z_2 - z_3)(\ln p - \ln p_1)(\ln p - \ln p_2)}{(\ln p_2 - \ln p_3)(\ln p_1 - \ln p_3)} \quad (1.5)$$

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  $\Lambda$ , from the model level above the lowest model level (subscript  $l - 1$ ), see the routine **CTSTAR**, using

$$T^* = T_{l-1} + \Lambda \frac{R_{\text{dry}}}{g} T_{l-1} \ln \frac{p_{\text{surf}}}{p_{l-1}} \quad (1.6)$$

$$T^* = \frac{\{T^* + \max[T_y, \min(T_x, T^*)]\}}{2} \quad (1.7)$$

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

$$T_0 = T^* + \Lambda \frac{\phi_{\text{surf}}}{g} \quad (1.8)$$

$$T_0 = \min[T_0, \max(T_x, T^*)] \quad (1.9)$$

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

$$\phi = \phi_{\text{surf}} - \frac{R_{\text{dry}} T^*}{\gamma} \left[ \left( \frac{p}{p_{\text{surf}}} \right)^\gamma - 1 \right] \quad (1.10)$$

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

**1.3.2 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 1.3.1](#)). Below the lowest model level wind components are assumed to be constant and equal to the values of the lowest model level.

**1.3.3 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 1.6.4](#).

(a) *Saturation vapour pressure*

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

$$e_{\text{sat}}(T) = a_1 \exp^{a_3 \left( \frac{T - T_3}{T - a_4} \right)} \quad (1.11)$$

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  $0^\circ\text{C}$  and the value over ice is taken for temperatures below  $-23^\circ\text{C}$ . For intermediate temperatures the saturation vapour pressure is computed as a combination of the values over water  $e_{\text{sat(water)}}$  and  $e_{\text{sat(ice)}}$  according to the formula

$$e_{\text{sat}}(T) = e_{\text{sat(ice)}}(T) + [e_{\text{sat(water)}}(T) - e_{\text{sat(ice)}}(T)] \left( \frac{T - T_i}{T_3 - T_i} \right)^2 \quad (1.12)$$

with  $T_3 - T_i = 23$  K.

(b) *Relative humidity*

In **GPRH** relative humidity  $U$  is computed from

$$U = \frac{pq \frac{R_{\text{vap}}}{R_{\text{dry}}}}{\left[ 1 + \left( \frac{R_{\text{vap}}}{R_{\text{dry}}} - 1 \right) q \right] e_{\text{sat}}(T)} \quad (1.13)$$

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 from 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

$$PWC_k = \frac{1}{g} \sum_{i=1}^k q_i (p_i - p_{i-1}) \quad (1.14)$$

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  $q = q_i$ .

(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.

### 1.3.4 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

$$T = \frac{(p_{\text{surf}} - p)T_l + (p - p_l)T^*}{p_{\text{surf}} - p_l} \quad (1.15)$$

Below the lowest model level the temperature is extrapolated by

$$T = T^* \left[ 1 + \alpha \ln \frac{p}{p_{\text{surf}}} + \frac{1}{2} \left( \alpha \ln \frac{p}{p_{\text{surf}}} \right)^2 + \frac{1}{6} \left( \alpha \ln \frac{p}{p_{\text{surf}}} \right)^3 \right] \quad (1.16)$$

with  $\alpha = \Lambda R_{\text{dry}}/g$ , for  $\phi_{\text{sat}}/g < 2000$  m, but  $\alpha$  is modified for high orography to  $\alpha = R_{\text{dry}}(T'_0 - T^*)/\phi_{\text{surf}}$ , where

$$T'_0 = \min(T_0, 298) \quad (1.17)$$

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

$$T'_0 = 0.002[(2500 - \phi_{\text{surf}}/g)T_0 + (\phi_{\text{surf}}/g - 2000) \min(T_0, 298)] \quad (1.18)$$

for  $2000 < \phi_{\text{surf}}/g < 2500$  m. If  $T'_0 < T^*$  then  $\alpha$  is reset to zero. The two temperatures  $T^*$  and  $T_0$  are computed using (1.6) to (1.9).

## 1.4 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 ATOVS (McNally *et al.*, 1999), AIRS, IASI, ATMS, as well as geostationary water vapour clear-sky radiances (Munro *et al.*, 2004). Clear, cloudy and precipitation-affected radiances from microwave imagers such as SSMI/S and microwave humidity sounders such as MHS are monitored or assimilated using an all-sky approach. A parallel datastream of AMSU-A data is also sent through the all-sky route for monitoring purposes, though AMSU-A is actively assimilated using the normal clear-sky route. Hence there are currently two observation operators in the IFS to assimilate satellite radiances, one for normal clear-sky radiances and totally overcast infrared radiances, and one for all-sky microwave 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 10.

### 1.4.1 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=NGTHR) or all-sky radiances (codetype=215=NSSMI).

The main set-up routine for radiances is **SURAD**. It recognises BUFR 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 platform and satellite numbers used by the RT-code (e.g. platform 1, satellite 10 for NOAA-10). New satellite IDs need to be defined in **GETSATID**, but the id-conversion tables can also 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 **YOMSATS**. The sensor number is used as index to various tables containing observation errors, BgQC thresholds, VarQC parameters, the  $J_0$ -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 **YOMSATS**). 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 that were found in the ODBs can be found in the *ifstraj* output by searching for **SATGRP**.

(c) *Radiative transfer coefficients, pressure levels and validation bounds*

Various preparations for RTTOV calculations are set-up in the call to **RTSETUP** from **SURAD**. This includes reading of the various coefficient files required by RTTOV. The files can be found under `/home/rd/rdx/data/38r1/sat/rttov`. There is one file with the prefix `rtcoef_` containing coefficients for the general clear-sky radiative transfer for each instrument and satellite, and files with prefix `scldcoef_` used for a parametrization of cloud scattering effects for some infrared sensors. Both types of files are read via a call to **RTTVI** (in the satrad library) which also communicates other settings from the RTTOV code in the satrad library to the IFS side. Only the files that are required are read, i.e. only the files for which observations are present (for the clear-sky coefficients) or for which data is present and the calculation of cloudy radiances is requested through switching the flags `LCLD_RTCALC_SCREEN` or `LCLD_RTCALC_ASSIM` on in the module **SATS\_MIX** for the respective sensor. A third type of RTTOV coefficient files (prefix `mietable_`) is required for RTTOV\_SCATT computations in the all-sky system (see 1.4.3), read via a call to **MWAVE\_SETUP** under **RTSETUP**.

If the interpolation to RTTOV pressure levels is performed on the IFS side (rather than inside RTTOV, see 1.4.2), the list of the `NLSAT = 44` fixed pressure levels is stored in arrays in **SATS\_MIX**. They are read in the satrad project from the transmittance coefficient file, called via the route of **SURAD** and **RTSETUP**. **RTSETUP** 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. 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. The reference pressure levels and limit profiles are only used if the interpolation to RTTOV pressure levels is performed in the IFS before the call to RTTOV instead of using the internal RTTOV interpolation (see 1.4.2), and the number of levels is fixed to `NLSAT = 44`.

If the interpolation to RTTOV pressure levels is performed inside RTTOV (the default behaviour at ECMWF, see 1.4.2), the transmittance coefficient file for different sensors can use any number of fixed pressure levels, and the number can be different for different sensors.

### 1.4.2 Clear-sky nadir radiances and overcast infrared nadir radiances

(a) *Radiance observation errors and bias correction*

Observation errors for 1C radiances are written to the ODB in a call to **RADICOB** (from **HRETR**). They are set by sensor and channel in the routine **DEFRUN**, variable `ROERR_RAD1C`.

The bias correction is performed through variational bias correction, see part II. The `VarBC` class is called “rad”, and the class-specific routines for the generic `VarBC` code are in the module **VARBC\_RAD**. For most sounding radiances, the predictors used consist of four layer-thicknesses derived from the First Guess (as defined in **VARBC\_PRED**), but some window channels do not include such airmass predictors to avoid aliasing of cloud information into the bias correction. Also, AMSU-A channel 14 is assimilated without a bias correction, in order to anchor the stratospheric temperature analysis. Without such an anchor, the variational bias corrections tend to drift to unrealistic values as a result of model biases. This is done at script level, through the namelist `NAMVARBC_RAD`.

(b) *Surface emissivity*

Microwave (**EMIS\_MW\_N**) and infrared (**EMIS\_IR**) surface emissivities are set during the screening phase in **RAD1CEMIS** (called from **HRETR**) and stored in the ODB for later use by RTTOV. Setting the emissivity to values outside the range of 0 and 1 prompts the calculation of surface emissivity within RTTOV, using **FASTEM** (Deblonde and English, 2001) for the microwave and **ISEM-6** (Sherlock, 1999) for the infrared. This is done for all microwave and infrared radiances over sea.

For microwave radiances over land, several options exist to specify the surface emission, following the methods described in Karbou *et al.* (2006). The surface emissivity can be specified through an atlas, or it can be dynamically retrieved from window channel observations and FG estimates of skin temperature and atmospheric profiles, or the skin temperature can be retrieved given an emissivity atlas value and FG estimates of atmospheric profiles. Default choices are made by sensor in **SUEMIS\_CONF**

(including which channel is used for the last two options) and can be overwritten through the namelist `NAMEMIS_CONF` or controlled through the preIFS switch `AMSU_LAND` in the Satellites window in the case of AMSU-A/B/MHS. For the two options with dynamic retrieval of emissivity or skin temperature, the required radiative transfer calculations are performed in the routine `satrat/rttov/ifs/rttov.ec` when called from `HRETR` via `RADTR` or `RADTR_ML` (see also next section). The atlas values and the retrieved emissivities or skin temperatures are written to the ODB, and used as fixed input in subsequent calls to `RTTOV`. If atlas values are required, these are read in the routine `DEFRUN`. The default for AMSU-A/B/MHS over land is to use the dynamic retrieval of surface emissivity, using an evolving emissivity atlas to quality-control the retrieved emissivities.

For some microwave sounders such as AMSU-A, a Kalman Filter is used to produce an evolving emissivity atlas from past dynamically retrieved emissivity values, as summarised in Krzeminski *et al.* (2009) and Bormann (2014). The atlas is updated using the programs `EMISKF_UPDATE` in the `satrad` project (together with `emiskf*` and `kfgrid*` routines that can be found in the `emiss` directory of the `satrad` project). The program accesses the ODB and reads the required retrieved emissivity values. Only emissivity values that have the datum status flag “use\_emiskf\_only” set during the blacklisting are considered. Routine `EMISKF_INIT` specifies the resolution of the atlas and other configuration settings. Routine `EMISKF_INIT_ATLAS` is used to read the atlas values, `EMISKF_TRAJ` to evaluate the new emissivity values against the atlas values, `EMISKF_PREDICT` to predict forward in time the evolution of the errors in the atlas emissivity, and `EMISKF_UPDATE_ATLAS` to perform the update of the emissivity parametrization using the Kalman Filter equations, and `EMISKF_WRITE_ATLAS` to output the updated atlas. To use the atlas with a new sensor/channel, include the new sensor/channel in the look-up table of known emissivity channels in `EMISKF_INIT`, and provide a new ODBsql view in `EMISKF_UPDATE`. The cycling of the atlas information is done through the files `emiskf.cycle*` which are stored as tar-ball in ECFS. The routine `EMISKF_INIT_ATLAS` is also used to read the atlas under `DEFRUN` during the application stage in the screening run. If no atlas is available a “coldstart” is performed, setting the atlas values and their errors to pre-specified values.

(c) *Calling the radiative transfer model*

The radiative transfer model `RTTOV` is called from the general observation operator routine `HOP`. More details on `RTTOV` can be found in Eyre (1991), updated by Saunders and Matricardi (1998).

`RTTOV` performs optical depth calculations on a number of fixed pressure levels, as specified in the `rtcoef_` coefficient file mentioned earlier. `RTTOV` includes the option to provide the atmospheric profile input either on this set of fixed pressure levels or on a set of different and variable pressure levels. In the latter case, `RTTOV` will perform the required interpolation internally, using an interpolation that provides smoother gradients than a simple linear interpolation. If this option is used, the radiative transfer computations are also performed on the input user levels, rather than the fixed `RTTOV` pressure levels.

The IFS allows the user to either perform the interpolation to the `RTTOV` pressure levels before the call to `RTTOV` on the IFS side, or to provide the model data on the model fields and use the `RTTOV` internal interpolation. The choice is done via the flag `LRTTOV_INTERPOL` in the module `SATS_MIX`. The default for ECMWF is to use the `RTTOV` internal interpolation, so the flag is set to `.True.`

Depending on the setting of `LRTTOV_INTERPOL` described in Hocking (2014), the routine `HOP` will perform different tasks:

In case `LRTTOV_INTERPOL` is set to `.False.`, `HOP` interpolates the model profiles of temperature, humidity and ozone ( $T$ ,  $q$ , and  $oz$ ) to the `NLSAT`(= 44) RT levels ( $\hat{T}$  and  $\hat{q}$ ) 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 are performed in the routine `HRADP`. For the purpose of radiance calculations  $T_{2m} = T_l$  (the lowest model level temperature) and  $q_{2m} = \hat{q}_{NLSAT}$  (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 1.6) in this context. In order to make the radiance cost function continuous in  $p_{surf}$  it was necessary to ensure that  $\hat{T}$  and  $\hat{q}$  approach  $T_{2m}$  and  $q_{2m}$  as the pressure on any of the RT levels approaches  $p_{surf}$ . This is

done in a section of **HRADP**. The routine **RADTR** checks that the input model profile is within the valid range of the transmittance regression, and re-sets the values to the limit values if they fall outside these validity bounds.

In case **LRTTOV\_INTERPOL** is set to **.True**. (the default), **HOP** will call **RTTOV** via the routine **RADTR\_ML** and profile information will be input to **RTTOV** on **NFLEVG+1** levels. These correspond to the **NFLEVG** model levels with an additional level set at the model's surface pressure, for which the temperature, humidity, and other gas concentration values are taken from the lowest model level. A check against the validity bounds of the **RTTOV** transmittance parametrization is in this case done within **RTTOV** (routine **SATRAD/RTTOV/MAIN/RTTOV\_CHECKINPUT.F90**), for the optical depth calculations only (switch **APPLY\_REGRESSION\_LIMITS** in **SATRAD/MODULE/RTTOV\_CONST.F90**). Various radiance preparations are again performed in a modified version of **HRADP**, **HRADP\_ML**.

In either case, the routine **HOP** constructs a list of requested channel numbers for each report from the observation array, and only model radiances for exactly those channels are then requested from the RT-code. The routines **RADTR** or **RADTR\_ML** packets the profiles into chunks of work of an appropriate maximum size for the RT-code (currently set to 8 in **SATRAD/MODULE/mod\_cparam.F90**). 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 the observation array are obtained just as it was done in the direct code. Consistency of the TL and adjoint operator can be tested by turning on the switch **LRTTOV\_ADTEST** in prepIFS.

#### (d) *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 skin temperature retrieval at the radiance FOVs is performed in the routine **HRADP/HRADPTL/HRADPAD**, called from **HOP/HOPTL/HOPAD**. The background skin temperature is provided by the model trajectory integration, and a background error of 1 K/5 K/7.5 K is assigned over sea/land/sea-ice, respectively (set in **SURAD**). The gradient with respect to the skin temperature obtained from **RTTOV** is temporarily stored in the **TOVSCVX** array and later transferred to its location in the distributed control vector. The next iteration of the minimisation provides updated skin temperature increments (also stored in **TOVSCVX**) that are used by **RTTOVTL** in subsequent iterations. The outer-loop iterations result in a new linearisation state, stored in **TOVSCVX5**. All the skin-temperature-related information at FOV locations that needs to be passed between job-steps, reside in the ODB, in the **skintemp** array. Here, **skintemp\_1** is the background skin temperature, and subsequent values are the values at the end of each following minimisation. The approach has been adopted for **CO<sub>2</sub>** retrieval at **AIRS** FOVs (Engelen *et al.*, 2004).

### 1.4.3 All-sky nadir radiances

Observations from microwave imagers and sounders can be assimilated using an all-sky approach which unifies clear-sky, cloudy, and precipitation-affected radiances in one observation operator, using **RTTOV-SCATT** for the radiative transfer, which is capable of modelling the effects of multiple scattering from hydrometeors. Three classes of microwave data are addressed by the all-sky route, with sometimes more than one class of data in the same instrument:

- (i) Microwave imagers **TMI** and some channels of **SSMIS** are actively assimilated and **WINDSAT**, **SSM/I** and **AMSR-2** are passively monitored, though only over oceans and only for latitudes equatorward of 60°
- (ii) Microwave humidity sounders such as **MHS** and some channels of **SSMIS** are actively assimilated over land, ocean and sea-ice.



- (iii) Microwave temperature sounders such as channels 1-5 of AMSU-A and some channels of SSMIS are monitored, but not assimilated. AMSU-A observations are assimilated actively through the clear-sky route.

All-sky observations follow a path through the IFS code that is slightly different from that for clear-sky radiance observations, though the observation operator does now run under **HOP** rather than under the model physics, as previously. However, the observation operator requires a number of diagnostic variables (e.g. precipitation flux and fraction) that come from the moist physics parametrizations and require a model timestep to have been run before they can be made available in observation space. These fields are generated in **CALLPAR**, stored in the **GFL** arrays and interpolated to observation space like any other model field required by an observation operator, using the **GOM** arrays. The main difference is that there is no interpolation: the observation operator gets the model profile at the grid point closest to the observation.

All-sky assimilation can be switched on in PrepIFS by means of switches for individual sensors, e.g. LSSMI, LAMSRE, LTMI, LSSMIS etc. For all-sky AMSU-A and MHS, which can run in parallel to clear-sky assimilation, there is a separate switch LAMSUA\_ALLKSY or LMHS\_ALLSKY. Comprehensive scientific documentation of the all-sky approach can be found in Bauer *et al.* (2010), Geer *et al.* (2010), Geer and Bauer (2011), Geer and Bauer (2012) and Geer *et al.* (2014).

#### (a) Preprocessing

Observations are received in BUFR format and are pre-processed by **SCRIPTS/GEN/premwing**, which calls a number of fortran programs:

**SATRAD/PROGRAMS/bufr\_screen\_ssmi\_1d** and **bufr\_screen\_amsre\_1d** perform a preliminary screening, removing observations over land and checking for any unrealistic brightness temperatures. AMSR-E observations are re-assigned to the BUFR subtype 127, which is that of SSM/I. Later, this is used to identify the observations as part of the all-sky path by giving them codetype 215 (see **ODB/CMA2ODB/buf2cmat\_new**).

**SATRAD/PROGRAMS/bufr\_grid\_screen** is called to do superobbing. Based on the final inner loop resolution, typically T255, observations are binned onto the Gaussian grid. This brings the relatively high observation resolution of microwave imagers (up to 10 km) down to roughly the scale of clouds in the IFS model. The superobbing is done by computing the numerical mean of all BUFR fields, except longitude and latitude, which are taken to be those of the grid-point, and the observation time, which comes from the last observation meaned. This program also allows a further thinning of the data by keeping only observations associated with grid points at every  $n$ th longitude and  $m$ th latitude.

#### (b) Observation operator

Code for the all-sky observation operator is prefixed by ‘mwave’ and located either in **IFS/MWAVE** or **SATRAD/MWAVE**. The observation operator wrapper **IFS/MWAVE/mwave\_wrapper** is called from **IFS/OP\_OBS/hop** or its TL or adjoint. The same wrapper function is called whether in direct, TL or AD mode, and the required TL or AD functionality is driven by optional arguments. Lower down, there are separate observation operator routines for each mode: (**IFS/MWAVE/mwave\_obsop**, **mwave\_obsop\_tl** or **mwave\_obsop\_ad**). In the screening trajectory, **IFS/MWAVE/mwave\_screen** is called instead. Inputs to the observation operator are the profiles of model variables (passed in via a structure of **mwave\_phys\_type**) and any observation-related information (passed via a structure of **mwave\_rad\_type**). The observation-related information has been read from the ODB by **IFS/MWAVE/mwave\_get**, **mwave\_get\_tl** or **mwave\_get\_ad**. Outputs from the observation operator (such as the simulated brightness temperatures, and the observation error) are returned in a similar way and are written to the ODB by **IFS/MWAVE/mwave\_put** or **mwave\_put\_tl**. This approach is slightly different to that of other observation operators and is a result of the code’s former location in the model physics.

The main function of the observation operator code in **IFS/MWAVE** and **SATRAD/MWAVE** is simply to provide the correct inputs and initialisations to run **RTTOV\_SCATT**. However, there is code for quality control, to produce diagnostic output, and to determine the observation error, which is not constant, but a function of hydrometeor amount, as described in Geer and Bauer (2011). VarQC (see part II), VarBC

(see part II), blacklisting and background quality control (all in [Chapter 3](#)) are done largely as for other radiance observations. The internal IFS thinning routines are completely bypassed, however, because a suitable thinning has already been achieved during the superobbing.

A number of initialisation tasks are performed in [IFS/MWAVE/mwave.setup](#), including reading the namelist [NAMMWAVE](#) for configuration flags. An array of `mwave_ids` structures is created, one for each satellite and sensor combination that will pass through the all-sky operators. In this private table are stored things like the instrument zenith angle, observation error specifications, and the ID numbers used in the rest of the IFS (e.g. sensor, satellite and bufr IDs).

(c) *External files*

Observation error definitions for completely clear and completely cloudy skies are stored in files with names like `mwave_error_<satellite>_<instrument>.dat`. These are read by [IFS/MWAVE/mwave.setup](#).

Some configuration options are specified in the [NAMMWAVE](#) namelist in [SCRIPTS/GEN/ifsmin](#) and [ifstraj](#).

(d) *Diagnostics*

The consistency of the TL and adjoint operators can be tested by setting `ldmwave_test = .true.` in the [NAMMWAVE](#) namelist in the [SCRIPTS/GEN/ifsmin](#) script. This causes [IFS/MWAVE/mwave\\_obsop\\_test](#) to be called during the minimisation with the real TL input values for each observation. The results of the test are written to the IFS logfiles, prefixed by 'MWAVE\_OBSOP\_TEST AD:'. The first number gives the TL / AD inaccuracy in multiples of machine precision. Typically this should be substantially less than 100 but it will go over 1000 for a few observations.

A number of diagnostics are stored in the ODB in the `allsky` or `allsky_body` table:

**DATUM\_TBFLAG** - This is a bitfield which records quality control decisions for the all-sky observations. It is an additional diagnostic on top of the usual status and event flags, and it only records decisions made internally in the all-sky observation operator. A value of 1 indicates an OK observation; all other values indicate rejection. However, even if the observation is considered OK by the all-sky observation operator, it may subsequently be rejected by the other IFS screening processes (e.g. blacklisting, thinning, VarQC, background QC), so always check the 'status@hdr' and 'status@body' too. Binary arithmetic can be used to decipher the `tbflag` bitfield. For example, if `DATUM_TBFLAG = 33 = 25 + 20`, that means bits 5 and 0 have been set. Bit 5 indicates contamination by sea-ice.  $2^0$  is equal to 1; this would have indicated "OK" if no other bit had been set. The full structure of the `DATUM_TBFLAG` bitfield is described in [IFS/MODULE/yommwave](#).

**DATUM\_TBCLOUD** - this is a bitfield recording the status of diagnostic cloud and rain identification tests performed on observed and simulated brightness temperatures by [IFS/OP\\_OBS/mwimager\\_cloud](#) and further documented in the code and in [Geer et al. \(2008\)](#). The bitfield structure is documented in [IFS/MODULE/yommwave](#). The lowest 2 bits give the results of the FG cloud test. With AND representing the bitwise boolean operator, `(DATUM_TBCLOUD AND 2) / 2` will give the result of the test for cloud in the FG, with 1 indicating a cloudy scene. `(DATUM_TBCLOUD AND 1) / 1` will give the result for the observation.

There are also a number of diagnostics relating to FG and analysis model state. These are valid at the time and location of the observation, giving information that is not otherwise archived. These values include the surface rain and snow rate, in  $\text{kg m}^{-2} \text{s}^{-1}$ , and the total columns of water vapour, cloud water, cloud ice, rain and snow, in  $\text{kg m}^{-2}$ .

#### 1.4.4 Cloud affected infrared radiances

For infrared data from HIRS, AIRS and IASI simplified cloud parameters (cloud top pressure and effective cloud fraction) are estimated for each field of view. Background values are computed during the screening in routine [CLOUD\\_ESTIMATE](#) using a method described in [McNally \(2009\)](#). If the scene is diagnosed as overcast (i.e. cloud fraction equal to 1) then all channels are used (that would be used in a completely



clear scene) and the cloud parameters become additional elements of the local control vector (as skin temperature). This is done by default, but can be disabled by setting the switch LCLDSINK to false in namelist NAMVAR. The cloud top pressure is assigned an error (CTOPBGE in YOMVAR currently equal to 5 hPa) but the cloud fraction is effectively fixed. The handling of the estimated cloud parameters is then performed in the routine HRADP/HRADPTL/HRADPAD, called from HOP/HOPTL/HOPAD. The gradient with respect to the cloud parameters is obtained from RTTOV (TOVSCVX array) and it is later transferred to its location in the distributed control vector. The next iteration of the minimisation provides updated cloud parameter increments (also stored in TOVSCVX) that are used by RTTOVTL in subsequent iterations. The outer-loop iterations result in a new linearisation state, stored in TOVSCVX5. All the cloud parameter information at FOV locations that needs to be passed between job-steps, resides in the ODB, in the satellite predictors table.

If the scene is not diagnosed as overcast, only channels flagged as clear are assimilated and the cloud parameters are essentially inactive.

The approach is also applied to SEVIRI all sky geostationary radiances to allow the additional assimilation of SEVIRI water-vapour overcast radiance observations in parallel with water-vapour clear sky radiances (Lupu and McNally, 2012).

#### 1.4.5 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 extension 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 two-dimensional facility is switched on by specifying `NOBSPROFS(10) > 1` for obstype 10 in the namelist `NAMNPROF`. Note that this means GPS radio occultation bending angles present in the assimilation will also take horizontal structure into account.

## 1.5 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).

### 1.5.1 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 `YOMSATS`.

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.*, 1998; Bormann *et al.*, 2003), much in the same way as conventional data, using the same interpolation routine. Other options are treating AMVs as layer averages, with weights defined, for instance, by a boxcar weighting function.

### 1.5.2 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 1.3), and the thickness is given by the difference  $\phi_T - \phi_B$ .

### 1.5.3 Gas retrievals

Retrievals of atmospheric species such as ozone or water vapour are used in the form of integrated layers bounded by a top and bottom pressure which are given as a part of the observation. The same observation operator is used as for precipitable water (`PPPWC`, Section 1.2). The same concept is applied to all data, whether it is total column data (like TOMS and GOME ozone data or MERIS total column water vapour) or data with higher vertical resolution (like SBUV). For ozone, variational bias correction is implemented (see Part II, class name “to3”, module `VARBC_TO3`). SBUV data is currently used as anchor for the variational bias correction of ozone and therefore assimilated without bias correction.

### 1.5.4 Scatterometer winds

In `HOP`, the observation array is scanned for SCAT data. For ERS-2 and ASCAT 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. Up to the first `NSCAWSOLMAX` (4 by default, adaptable through the namelist `NAMSCC`) wind solutions are accepted. In case only one ambiguity is found, the report is 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`). For winds that are not closest to the first guess or analysis, global datum event flag 9, respectively 10 is set (see Table 2.29). For the latter case datum status is set to rejected as well (Table 2.28). When `LQSCATT = .FALSE.`, the two first winds are used and the ambiguity removal takes place implicitly through a special SCAT cost-function (see part II), in `HJO` (Stoffelen and Anderson, 1997). In that case for QuikSCAT the most likely wind (highest *a priori* probability) and its most opposing ambiguity are selected.

Routine `PPUV10M` (Section 1.6) is like `SYNOP`, `SHIP` and `DRIBU` wind, also used also for SCAT data. Difference is that (in case account is taken for ocean current) the relative wind rather than the absolute

wind is returned (not active in the operational suite, though), and the evaluation of equivalent-neutral wind rather than the real wind (which latter includes for scatterometer data undesired stability effects; operational since November 2011).

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

### 1.5.5 GPS Radio Occultation bending angles

The subroutine **GPSRO\_OP** is called in **HOP** and it simulates GPS radio occultation bending angles using the one-dimensional model outlined in [Healy and Thépaut \(2006\)](#). The subroutine evaluates a profile of bending angles,  $\alpha$  as function of impact parameter,  $a$ , at each observation location by evaluating the integral

$$\alpha(a) = -2a \int_a^\infty \frac{\frac{d \ln n}{dx}}{(x^2 - a^2)^{1/2}} dx \quad (1.19)$$

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 1.3.1](#)) 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.

## 1.6 SURFACE OBSERVATION OPERATORS

Preparations for the vertical interpolation of surface data is done in **PREINTS** (see [Subsection 1.1.4](#)). Here dry static energy (**SURBOUND**), Richardson number, drag coefficients and stability functions (**EXCHCO**) are computed. For scatterometer data, information on equivalent neutral 10-metre wind is directly fetched from the physics package (via the GOM arrays) which is preprocessed in routine **EXCHCO\_VDF**. The actual vertical interpolation is performed in **PPOBSAS**, which embraces routines for 10-metre vector-wind components (**PPUV10M**), 2-metre temperature (**PPT2M**) and 2-metre relative humidity (**PPRH2M**).

### 1.6.1 Vertical interpolation

For wind and temperature an analytical technique ([Geleyn, 1988](#)) is used to interpolate values between the lowest model level and the surface. It is based on Monin–Obukhov theory in which simplified versions of stability functions  $\phi_M$  and  $\phi_H$  are used. The following equations are to be integrated:

$$\frac{\partial \mathbf{u}}{\partial z} = \frac{\mathbf{u}_*}{\kappa(z + z_0)} \phi_M \left( \frac{z + z_0}{L} \right), \quad (1.20)$$

$$\frac{\partial s}{\partial z} = \frac{s_*}{\kappa(z + z_0)} \phi_H \left( \frac{z + z_0}{L} \right), \quad (1.21)$$

$$L = \frac{c_p T u_*^2}{g \kappa s_*}, \quad (1.22)$$

were  $\mathbf{u}$ ,  $s$  are wind and energy variables,  $\mathbf{u}_*$ ,  $s_*$  are friction values,  $u_* = |\mathbf{u}_*|$ , and  $\kappa = 0.4$  is von Kármán's constant. Note that  $\mathbf{u}$  denotes the vector wind relative to a surface current  $\mathbf{u}_0$ ,

$$\mathbf{u} = \mathbf{u}_a - \mathbf{u}_0, \quad (1.23)$$

with  $\mathbf{u}_a$  the wind in the absolute (model) frame. In default configuration (global variable LECURR is false) surface current is zero, in which case the distinction between absolute and relative wind is irrelevant.

The temperature is linked to the dry static energy  $s$  by

$$s = c_p T + \phi \quad (1.24)$$

$$c_p = c_{p_{\text{dry}}} \left[ 1 + \left( \frac{c_{p_{\text{vap}}}}{c_{p_{\text{dry}}}} - 1 \right) q \right]. \quad (1.25)$$

The neutral surface exchange coefficient at the height  $z$  is defined as

$$C_N = \left[ \frac{\kappa}{\ln\left(\frac{z+z_0}{z_0}\right)} \right]^2, \quad (1.26)$$

where  $z_0$  is the surface roughness length. Drag and heat coefficients are defined as

$$C_M = \frac{u_*^2}{[u(z)]^2}, \quad (1.27)$$

$$C_H = \frac{u_* s_*}{u(z)[s(z) - \tilde{s}]}, \quad (1.28)$$

where  $u(z) = |\mathbf{u}(z)|$  and  $\tilde{s}$  is the dry static energy at the surface. Details on the estimation of the roughness length and transfer coefficients can be found in [Subsection 1.6.3](#).

For convenience the following quantities are introduced:

$$B_N = \frac{\kappa}{\sqrt{C_N}}, \quad B_M = \frac{\kappa}{\sqrt{C_M}}, \quad B_H = \frac{\kappa \sqrt{C_M}}{C_H}. \quad (1.29)$$

For stable conditions the (simplified) stability function is assumed

$$\phi_{M/H} = 1 + \beta_{M/H} \frac{z}{L}, \quad (1.30)$$

and integration of (1.20) and (1.21) from 0 to  $z_1$  (the lowest model level) leads to values for relative wind  $\mathbf{u}(z)$  and static energy  $s(z)$  at observation height  $z$ :

$$\mathbf{u}(z) = \frac{\mathbf{u}(z_1)}{B_M} \left[ \ln\left(1 + \frac{z}{z_1}(e^{B_N} - 1)\right) - \frac{z}{z_1}(B_N - B_M) \right], \quad (1.31)$$

$$s(z) = \tilde{s} + \frac{s(z_1) - \tilde{s}}{B_H} \left[ \ln\left(1 + \frac{z}{z_1}(e^{B_N} - 1)\right) - \frac{z}{z_1}(B_N - B_H) \right]. \quad (1.32)$$

In unstable conditions the stability function is chosen as

$$\phi_{M/H} = \left(1 - \beta_{M/H} \frac{z}{L}\right)^{-1} \quad (1.33)$$

and the vertical profiles for relative wind and dry static energy are then given by

$$\mathbf{u}(z) = \frac{\mathbf{u}(z_1)}{B_M} \left[ \ln\left(1 + \frac{z}{z_1}(e^{B_N} - 1)\right) - \ln\left(1 + \frac{z}{z_1}(e^{B_N - B_M} - 1)\right) \right], \quad (1.34)$$

$$s(z) = \tilde{s} + \frac{s(z_1) - \tilde{s}}{B_H} \left[ \ln\left(1 + \frac{z}{z_1}(e^{B_N} - 1)\right) - \ln\left(1 + \frac{z}{z_1}(e^{B_N - B_H} - 1)\right) \right]. \quad (1.35)$$

In case the influence of stability is neglected, the following equivalent-neutral wind profile  $\mathbf{u}_n(z)$  is obtained:

$$\mathbf{u}_n(z) = \frac{\mathbf{u}(z_1)}{B_M} \ln \left( 1 + \frac{z}{z_1} (e^{B_N} - 1) \right). \quad (1.36)$$

For wind, the relevant routine **PPUV10M** embodies this method of [Geleyn \(1988\)](#) to estimate vector wind components at observation height  $z$  from provided lowest model level wind  $\mathbf{u}(z_1) = \mathbf{u}_a(z_1) - \mathbf{u}_0$ . For scatterometer data, by default, relative wind (1.31), (1.34) is returned, while for all other data the wind in the absolute frame is evaluated:

$$\mathbf{u}_a = \mathbf{u} + \mathbf{u}_0. \quad (1.37)$$

For scatterometer data, by default equivalent-neutral wind (1.36) is returned. In case non-neutral wind is to be assimilated (operational before November 2010), a variable `LSCATT_NEUTRAL` is to be set to false.

The temperature at observation height  $z = 2$  m is evaluated in **PPT2M**. It is obtained from  $s$  as

$$T(z) = s(z) - \frac{zg}{c_p}, \quad (1.38)$$

where  $s$  is interpolated according to (1.32) and (1.35).

The vertical interpolation relies on estimates for coefficients  $B_M, B_N$  for wind, and on coefficients  $B_H, B_N$  and dry static energy at the surface  $\tilde{s} = \tilde{s}(T_{\text{surf}}, q = 0)$ . These are provided in the routines **EXCHCO**, **EXCHCO\_VDF** and **SURBOUND**, and are described in the following two subsections.

### 1.6.2 Surface values of dry static energy

To determine the dry static energy at the surface we use (1.24) and (1.25) where the humidity at the surface is defined by

$$\tilde{q} = q(z = 0) = h(C_{\text{snow}}, C_{\text{liq}}, C_{\text{veg}}) q_{\text{sat}}(T_{\text{surf}}, p_{\text{surf}}) \quad (1.39)$$

where, according to [Blondin \(1991\)](#),  $h$  is given by

$$h = C_{\text{snow}} + (1 - C_{\text{snow}})[C_{\text{liq}} + (1 - C_{\text{liq}})\bar{h}] \quad (1.40)$$

with

$$\bar{h} = \max \left\{ 0.5 \left( 1 - \cos \frac{\pi \vartheta_{\text{soil}}}{\vartheta_{\text{cap}}} \right), \min \left( 1, \frac{q}{q_{\text{sat}}(T_{\text{surf}}, p_{\text{surf}})} \right) \right\} \quad (1.41)$$

where  $\vartheta_{\text{soil}}$  is the soil moisture content and  $\vartheta_{\text{cap}}$  is the soil moisture at field capacity (2/7 in volumetric units). Equation (1.40) 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  $C_{\text{snow}}$  depends on the snow amount  $W_{\text{snow}}$  so that

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

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

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

where

$$W_{\text{liq}_{\text{max}}} = W_{\text{layer}_{\text{max}}} \{ (1 - C_{\text{veg}}) + C_{\text{veg}} A_{\text{leaf}} \}$$

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

### 1.6.3 Transfer coefficients

Comparing the (1.20) and (1.21) integrated from  $z_o$  to  $z + z_o$  with (1.26) to (1.28),  $C_M$  and  $C_H$  can be analytically defined:

$$\frac{1}{C_M} = \frac{1}{\kappa^2} \left[ \int_{z_o}^{(z+z_o)} \frac{\phi_M(z'/L)}{z'} dz' \right]^2 \quad (1.42)$$

$$\frac{1}{C_H} = \frac{1}{\kappa^2} \left[ \int_{z_o}^{(z+z_o)} \frac{\phi_M(z'/L)}{z'} dz' \int_{z_o}^{(z+z_o)} \frac{\phi_H(z'/L)}{z'} dz' \right] \quad (1.43)$$

Because of the complicated form of the stability functions, the former integrals have been approximated by analytical expressions, formally given by (coded in **EXCHCO**)

$$\begin{aligned} C_M &= C_N f_M \left( Ri, \frac{z}{z_o} \right) \\ C_H &= C_N f_H \left( Ri, \frac{z}{z_o} \right) \end{aligned} \quad (1.44)$$

where  $C_N$  is given by (1.26). The bulk Richardson number  $Ri$  is defined as

$$Ri = \frac{g \Delta z \Delta T_v}{c_p T_v |\Delta u|^2} \quad (1.45)$$

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*  $Ri < 0$

$$f_M = 1 - \frac{2bRi}{1 + 3bcC_N \sqrt{\left(1 + \frac{z}{z_o}\right)(-Ri)}}, \quad (1.46)$$

$$f_H = 1 - \frac{3bRi}{1 + 3bcC_N \sqrt{\left(1 + \frac{z}{z_o}\right)(-Ri)}}, \quad (1.47)$$

with  $b = c = 5$ .

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

$$f_M = \frac{1}{1 + 2bRi / \sqrt{(1 + dRi)}}, \quad (1.48)$$

$$f_H = \frac{1}{1 + 3bRi / \sqrt{(1 + dRi)}}, \quad (1.49)$$

with  $d = 5$ .

(a) *Extraction of stability information from the ECMWF surface-layer physics*

The estimation of transfer coefficients as described above (Louis, 1979; Louis *et al.*, 1982) does not overlap well with the stability as evaluated in the full nonlinear surface layer physics parametrization, for two reasons. First, the method of Louis (1979); Louis *et al.* (1982) does not correspond anymore with the presently used parametrization. And second, the estimation of the neutral exchange coefficient (1.26) uses (for technical reasons) a roughness length  $z_o$  that is based on climatology, rather than on the actual roughness. Over oceans this embraces a value of  $z_o = 1\text{mm}$ , which is typically one order of magnitude too high. The effect on the estimation on 10-metre wind appears to be negligible, however, for 10-metre equivalent neutral wind (used for scatterometer data) and wind at 4 or 5 metre height (typical buoy observation height) a systematic effect can be observed (Hersbach, 2010a).

As an alternative, the information on stability can be extracted from the 10-metre equivalent neutral wind  $u_n$  as evaluated in `SPPCFL_MOD` in the ECMWF surface-layer physics, which is activated by a switch `LVDFTRAJ=true`. In that case, over the ocean roughness length  $z_0$  is estimated from  $u_n$  and the ocean-wave Charnock parameter  $\alpha$ , by an approximate solution (Hersbach, 2011) of the following set of implicit equations (coded in `ZOSEA`):

$$z_0 = \alpha_M \frac{\nu}{u_*} + \alpha \frac{u_*^2}{g}, \quad (1.50)$$

$$u_n = \frac{u_*}{\kappa} \log(1 + z_{10}/z_0). \quad (1.51)$$

Here  $z_{10} = 10$  m,  $\alpha_M = 0.11$ ,  $\kappa = 0.4$  is the Von Kármán constant,  $g = 9.80665 \text{ ms}^{-2}$  is the gravitational acceleration, and  $\nu = 1.5 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$  the kinematic viscosity.

The coefficient  $C_N$  is, again evaluated by (1.26), but now using the improved estimate of  $z_o$ , while coefficient  $C_M$  for momentum is given by (`EXCHCO_VDF`):

$$C_M = B_M^2 / \kappa, \quad B_M = \log(1 + z_{10}/z_0) \|\mathbf{u}(z_1)\| / u_n. \quad (1.52)$$

In the operational configuration this method is only used for the assimilation of scatterometer wind. For other observables, the method of Louis (1979); Louis *et al.* (1982) (`LVDFTRAJ=false`) is used. Routine `EXCHCO_VDF` does not provide an estimate for the coefficient  $C_H$  for heat.

#### 1.6.4 Two-metre relative humidity

In `GPRH` relative humidity is computed according to (1.13). 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 2

# Observation processing

### Table of contents

<b>2.1</b>	<b>Basic principles</b>
<b>2.2</b>	<b>Main MKCMARPL tasks and functions</b>
2.2.1	Basic observation processing setup
2.2.2	Invoking, initializing and controlling the MKCMARPL
2.2.3	MKCMARPL <i>[DEPRECATED]</i>
2.2.4	Basic observation handling routines <i>[DEPRECATED]</i>
<b>2.3</b>	<b>Migration towards Continuous Observation Processing Environment (COPE)</b>
<b>2.4</b>	<b>Observation types, subtypes and code types</b>
2.4.1	BUFR observation types and subtypes
2.4.2	ODB observation and code types
2.4.3	Mapping between ODB and BUFR observation types, code types and subtypes
<b>2.5</b>	<b>Variables</b>
2.5.1	Observed variables
2.5.2	Derived variables
2.5.3	Adjusted variables
2.5.4	Codes for variables
<b>2.6</b>	<b>Observation error statistics</b>
2.6.1	Persistence observation error
2.6.2	Prescribed observation errors
2.6.3	Derived observation errors
2.6.4	Final (combined) observation error
<b>2.7</b>	<b>Definitions</b>
2.7.1	Observation characteristics: instrument specification and retrieval type
2.7.2	Vertical coordinate: pressure, satellite ID and level ID codes
2.7.3	ODB report status: events, flags and codes
2.7.4	Datum status: events, RDB and analysis flags

## 2.1 BASIC PRINCIPLES

The ECMWF Data Assimilation Observation Processing System is split into two parts.

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

In general, observation processing that is independent of the model background is carried out in the Non-IFS related modules.

The non-IFS observation processing has two main sections.

- (i) Preparation and massaging of input BUFR data.
- (ii) Creation of Observation Database (ODB) files ready for use inside IFS.

The first part of the non-IFS observation processing 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.

The second part of the non-IFS observation processing, **BUFR2ODB**, extracts a subset of the BUFR data into new "ECMA" ODB files.

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.

## 2.2 MAIN MKCMARPL TASKS AND FUNCTIONS

### 2.2.1 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. 2.1](#)).

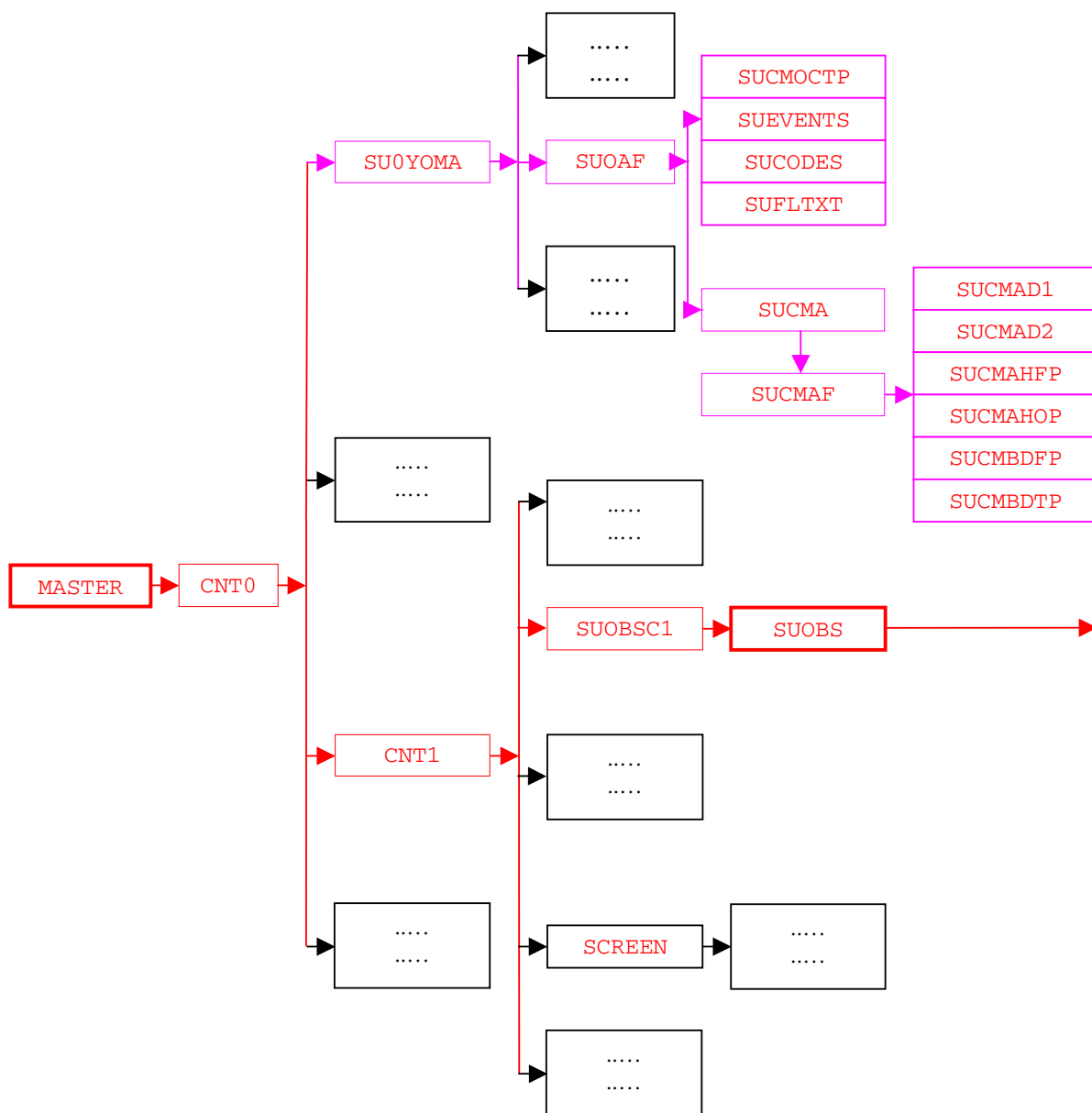
- Program **MASTER** calls **CNT0** which in turn calls **SU0YOMA**.
- **SU0YOMA** calls (among other routines) **SUOAF** from which **SUCMOCTP**, **SUEVENTS**, **SUCODES**, **SUFLTXT** and **SUCMA** are called. **SUCMOCTP** defines the ODB observation types and code types, and **SUEVENTS**, **SUCODES** 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 (DDR) as well as the ODB packing patterns (bit structure) employed for header and body respectively.

### 2.2.2 Invoking, initializing and controlling the MKCMARPL

The **MKCMARPL** run is initiated by the **MKCMARPL** subroutine (see [Fig. 2.2](#)). 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**.

- **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** deals with initialising **SCATT** 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



**Figure 2.1** 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. 2.2* continues the flow diagram from *SUOBS*.

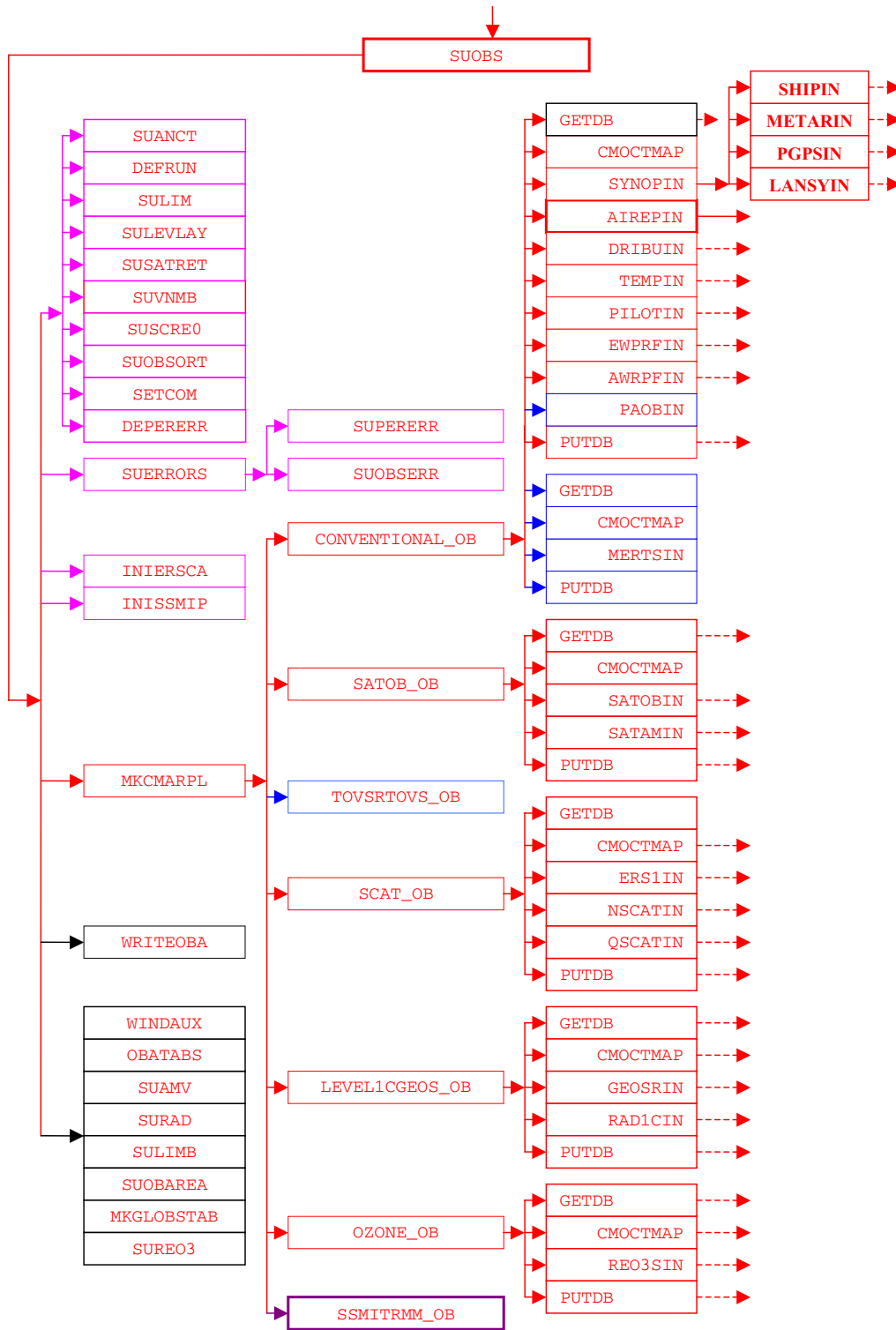


Figure 2.2 (Continued from Fig. 2.1) Simplified IFS observation pre-processing flow diagram (SUOBS). Colour coding scheme: (a) routines in red boxes perform observation pre-processing, (b) routines in pink boxes carry out observation pre-processing set up, (c) routines in black boxes are not directly involved in observation pre-processing, (d) routines in blue boxes are obsolete and (e) routines in plum boxes are awaiting revision. Fig. 2.3 continues the flow diagram from AIREPIN.

dealt with by calling **MKCMARPL** (more about it in [Subsection 2.2.3](#)). After **MKCMARPL** has finished there are several ways to proceed. These depend on the status of **LMKCMARPLO** and **LRPLSWAPOUT** 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.
```

### 2.2.3 **MKCMARPL** [DEPRECATED]

**WARNING:** As of IFS cycle 41r1, majority of **MKCMARPL** tasks dealing with conventional observation pre-processing have been deprecated and their functionality has been migrated to **COPE** framework. Although the **MKCMARPL** has not been fully replaced yet, and can still be used, it is strongly recommended to do new developments in the area of conventional observations within the **COPE** framework. For more details, including a full list of deprecated subroutines, see [Section 2.3](#).

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 six groups: **CONVENTIONAL**, **SATOB**, **TOVS/RTOVS**, **SCATT**, **LEVEL1C/GEOSS** and **OZONE** observations. For each group a separate subroutine is called: **CONVENTIONAL\_OB**, **SATOB\_OB**, **TOVSRTOVS\_OB**, **SCAT\_OB**, **LEVEL1CGOES\_OB** and **OZONE\_OB**. These routines are just cover or hat routines for the actual work to be carried out underneath. However, **TOVSRTOVS\_OB** is currently not called because it is obsolete.

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: **SYNOBIN**, **AIREPIN**, **DRIBUIN**, **TEMPIN**, **PILOTIN**, **EWPRFIN**, **AWPRFIN**, **PAOBIN** and **METARSIN**. A worker routine name indicates which observations it is dealing with.
- **SATOB\_OB** calls **SATOBIN** and **SATAMIN**.
- **SCAT\_OB** calls **ERS1IN**, **NSCATIN**, **ASCATIN** and **QSCATIN**.
- **LEVEL1CGEOS\_OB** calls **RAD1CIN** and **GOESRIN**.
- **OZONE\_OB** calls only **REO3SIN**.

### 2.2.4 **Basic observation handling routines** [DEPRECATED]

The observation pre-processing worker routines referred to in [Subsection 2.2.3](#), 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. 2.3](#)).

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**).

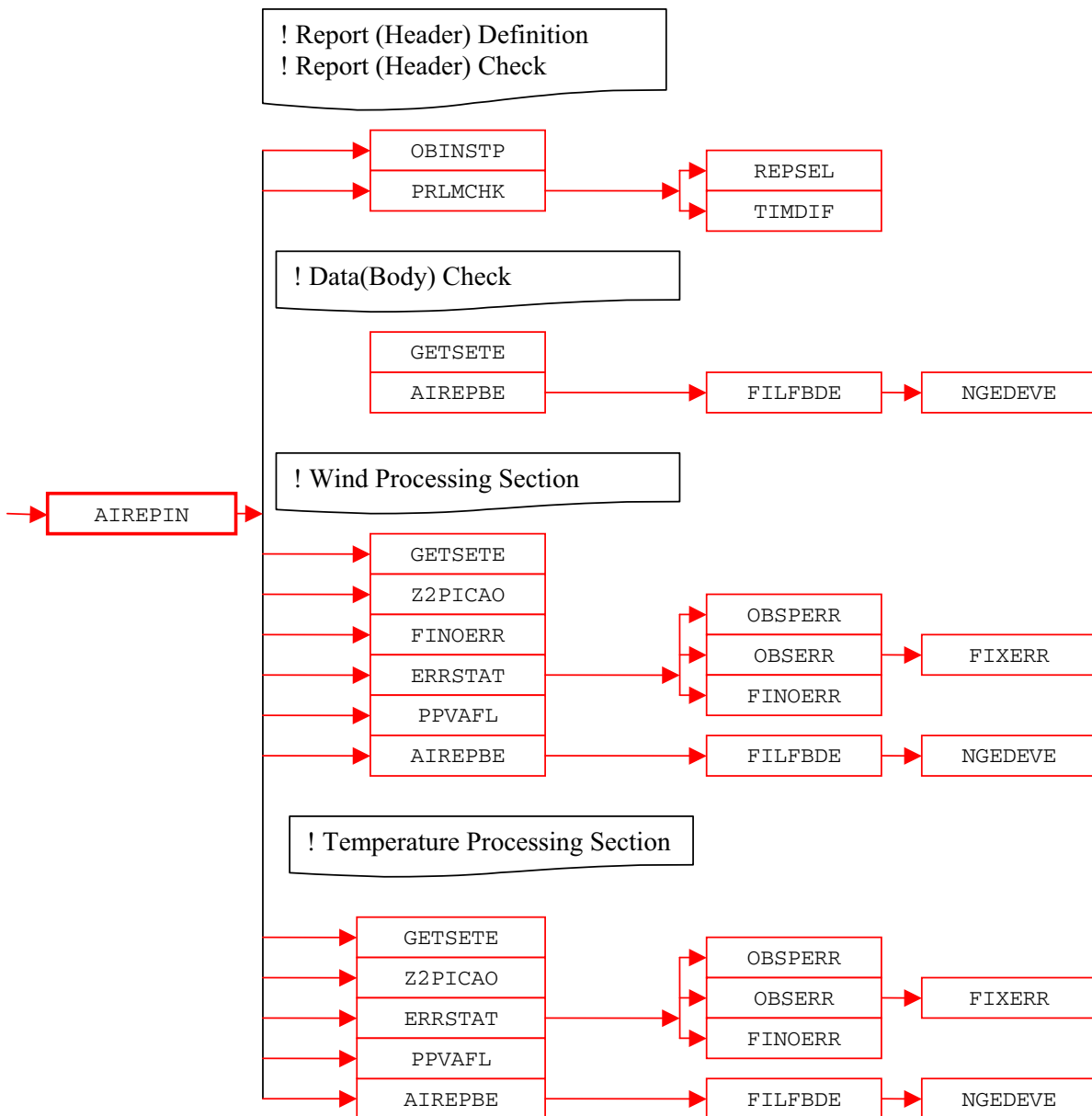


Figure 2.3 (Continued from Fig. 2.2) Simplified IFS observation pre-processing flow diagram (AIREPIN). Colour coding scheme: routines in red boxes perform observation pre-processing.

- **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.

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 (*SU\_ERRORS*).

The pattern of activities outlined for *AIREPIN* is repeated more or less in the other worker routines. However, the *SYNOBIN* routine is first split further into *SHIPIN*, *METARIN*, *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.

## 2.3 MIGRATION TOWARDS CONTINUOUS OBSERVATION PROCESSING ENVIRONMENT (COPE)

The main objective of COPE project is to consolidate various observation processing tasks that are either carried out in IFS screening, BUFR2ODB conversion or in preobs tasks, into a unified modular framework. The first phase of the project was focused on tasks that do not require model information and thus can be externalized from IFS. Decoupling observation pre-processing from the actual assimilation run has many benefits. Perhaps the most important being to free computational resources in time critical path and to increase resilience against anomalies in observing systems.

Sofar, only conventional observation pre-processing has been fully externalized from IFS, replacing the functionality of *MKCMARPL*. To activate COPE framework, one needs to set *LCOPE* option to *true* in PrepIFS (this is default since 40r3). Activating COPE will automatically disable *MKCMARPL* worker subroutines for all conventional observations (e.g. *AIREPIN*, *METARIN*, etc.). This is done in ifstraj script using *LMKCMARPL* logical array in *NAMOBS* namelist to deactivate processing for selected observation types.

The following is an alphabetical list of all subroutines, modules and module variables that have been deprecated and that are planned to be removed from IFS in one of the future cycles:

*AIREPBE*, *AIREPIN*, *AWPRFIN*, *BIASCOR\_ODB*, *CONVENTIONAL\_OB*, *DRIBUBE*, *DRIBUIN*, *ERRSTAT*, *EWPRFIN*, *FINOEREV*, *FIXERR*, *GET\_NEW\_RS\_TRH\_BIAS*, *GET\_RS\_T\_BIAS*, *HATBIASC*, *LNSYIN*, *METARIN*, *NEW\_RS\_TRH\_BIAS*, *OBINSTP*, *OBSERR*, *PGPSIN*, *PILOTBE*, *PILOTIN*, *PRLMCHK*, *PTENDCOR*, *REPSEL*, *RH2Q*, *RS\_BIAS\_VALIDITY*, *SHIPIN*, *SONDE\_COUNTRY\_DB\_MATCH*, *SULEVLAY*, *SUOBSERR*, *SYNOBPE*, *SYNOBIN*, *TEMPBE*, *TEMPIN*, *YOMLVLY*, *YOMMKODB:LCD[0-9]\**, *YOMMKODB:NMKCMVSE*, *YOMOERR*, *YOMOBS:LMKCMARPL*, *YOMRSTRHBIAS*, *Z2PICAO*.

Additionally, some of the subroutines considered to be completely obsolete were not reimplemented in COPE framework. These include **PTENDCOR** formerly used for ship surface pressure adjustment (Subsection 2.5.3) and subroutines **GET\_RS\_T\_BIAS**, **HATBIASC**, **BIASCOR\_ODB**, **RS\_BIAS\_VALIDITY** and **BIASCOR** used for old style radiosonde temperature and humidity bias correction. These subroutines are not likely to be reimplemented in COPE unless there is a specific need to do so.

The following listing should give some pointers as to where to find particular **MKCMARPL** functionalities within the COPE framework. On the left side (red font) are former IFS Fortran subroutines and, on the right (bold font), their corresponding counterparts in COPE framework. Note that this is only represents a loose mapping where applicable:

- **AIREPIN** — **airep.json**
- **DRIBUIN** — **dribu.json**
- **TEMPIN** — **temp.json**
- **PILOTIN** — **pilot.json**
- **PGPSIN** — **pgps.json**
- **LNSYIN** — **synop.json**
- **SHIPIN** — **ship.json**
- **AWPRFIN** — **awp.json**
- **EWPRFIN** — **ewp.json**
- **PRLMCHK** — **DateTimeValidator, LocationValidator**
- **YOMOERR, SUOBSERR** — **error\_statistics.csv**
- **YOMLVLY, SULEVLAY, ERRSTAT, OBSERR, FIXERR** — **PrescribedErrorAssigner**
- **FINOEREV** — **FinalErrorAssigner**
- **RH2Q** — **SpecificHumidityAssigner**
- **OBINSTP** — **InstrumentTypeAssigner**
- **Z2PICA0** — **HeightToPressureConverter**
- **YOMRSTRHBIAS, GET\_NEW\_RS\_TRH\_BIAS, SONDE\_COUNTRY\_DB\_MATCH, NEW\_RS\_TRH\_BIAS** — **RadiosondeBiasCorrector**

As can be seen in the listing above, there is nearly one to one correspondence between the former **MKCMARPL** worker Fortran subroutines **[A-Z]\*IN** and **[a-z]\*.json** configuration files. This is not by coincidence since preserving the former logical structure makes the transition to new framework more transparent. The JSON configuration files describe processing pipeline for the given observation type in declarative way. This allows to modify or even construct new pipelines at runtime, reusing existing filters, without having to recompile the source. Conceptually, every pipeline is composed of a chain of filters that are sequentially applied on each observation report. All JSON configuration files can be found in the standard IFS scripts directory.

Since COPE is a collaborative project involving several external partners, its source code is managed under common Git repository: <https://software.ecmwf.int/stash/projects/COPE/repos/cope>.

## 2.4 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).

### 2.4.1 BUFR observation types and subtypes

Some example BUFR observation types and subtypes are listed in Table 2.1. See WMO documentation for the full set, or the ODB governance tables <http://data-portal.ecmwf.int/odbgov/>



**Table 2.1** Some BUFR observation types and subtypes. See WMO documentation for the full set, or the ODB governance tables <http://data-portal.ecmwf.int/odbgov/>.

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
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 2.1 *Continued.*

Observation Type		Subtype	
Code	Name	Code	Name
		87	AMV
		89	Geostationary Clear Sky Radiances (GRAD)
		189	Geostationary Clear Sky Radiances (GRAD)
		190	Geostationary All Sky Radiances (GRAD)
12	SCATT/SSMI	122	ERS-1, ERS-2
		127	SSMI
		136	NSCAT
		137	QSCAT
		139	ASCAT
253	PAOB	161	PAOB

### 2.4.2 ODB observation and code types

There are also ODB ‘observation types’ and, as with BUFR, there are different number of code types for each of them. It is 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. Some ODB observation types and code types are listed in Table 2.2. See the ODB governance tables <http://data-portal.ecmwf.int/odbgov/> for a fuller and more up-to-date list.

### 2.4.3 Mapping between ODB and BUFR observation types, code types and subtypes

As indicated in Subsection 2.4.2 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. See the ODB governance tables <http://data-portal.ecmwf.int/odbgov/> for the latest information.

## 2.5 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 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.

### 2.5.1 Observed variables

The exact list of what is observed or present in the list of BUFR observation types and sub types is long and rapidly changing. Therefore Table 2.3 lists just those variables associated with conventional observations.

### 2.5.2 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.

**Table 2.2** *Some ODB observation types and code types. See the ODB governance tables <http://data-portal.ecmwf.int/odbgov/> for a fuller and more up-to-date list*

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
141	AIREP		
142	Simulated AIREP		
144	AMDAR		
145	ACARS		
241	COLBA		
3	SATOBS	88	SATOBS
		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 2.2 *Continued.*

Observation Type		Code Type	
Code	Name	Code	Name
8	PAOB	164	PAOB
9	SCATTEROMETER	122	ERS-1, ERS-2
		210	NSCAT
		301	QuikSCAT
		139	ASCAT
10	RAW RADIANCE		

 Table 2.3 *Observed variables for conventional observations.*

Observation Type		Observed Variable
BUFR	ODB	
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.
AIREP	AIREP	Upper Air Wind Direction/Force ( $DDD/FFF$ ) Temperature ( $T$ )

- (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 ( $\sigma^0$ 's) are transformed into several pairs of ambiguous wind components ( $u$  and  $v$ ); this actually involves a retrieval according to some model function describing the relationship between winds and  $\sigma^0$ 's and requires a fair bit of computational work. Details of this procedure are given in [Section 3.5](#).
- (iv) Mean layer temperature is transformed into thickness ( $DZ$ ) for SATEM and TOVS observations.

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

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

The  $RH$  is derived using

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

where function  $F$  of either  $T$  or  $Td$  is expressed as

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

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.

### 2.5.3 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

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

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,  $\theta$  is the latitude and  $\Omega = 0.7292 \times 10^{-4} \text{ s}^{-1}$  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  $Z = 0.11 \text{ kgm}^{-3}$  is an assumed air density. Now the adjusted pressure tendency ( $P_t^a$ ) is found as

$$P_t^a = P_t - \left( U_s \frac{\partial p}{\partial x} + V_s \frac{\partial p}{\partial y} \right)$$

Finally, the adjusted surface pressure ( $P_s^a$ ) is found as

$$P_s^a = P_s - P_t^a \Delta t$$

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

#### 2.5.4 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. For illustrative purposes some historical codes are listed in [Table 2.4](#) but for the latest information please check the subroutine **SUVNMB** or ideally the ODB governance database <http://data-portal.ecmwf.int/odbgov/>.

## 2.6 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.

### 2.6.1 Persistence observation error

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

- (i) Season.
- (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  $\theta$  is latitude.

The persistence error for time difference between analysis and observation  $\Delta t$  is then expressed as a function of  $b$  with a further dependence on latitude and a maximum persistence error  $E_{\max\text{pers}}$  for 24 hour given by

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

where  $\Delta t$  is expressed as a fraction of a day. The  $E_{\max\text{pers}}$  have the values shown in [Table 2.5](#).

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**.

**Table 2.4** Some variables (VARNO) in the ODB. Check the ODB governance database <http://data-portal.ecmwf.int/odbgov/> for the latest information.

No.	Code	Name	Unit
1	3	Wind Component ( $u$ )	$\text{ms}^{-1}$
2	4	Wind Component ( $v$ )	$\text{ms}^{-1}$
3	1	Geopotential ( $Z$ )	$\text{m}^2\text{s}^{-2}$
4	57	Thickness ( $DZ$ )	$\text{m}^2\text{s}^{-2}$
5	29	Relative Humidity ( $RH$ )	numeric
6	9	Precipitable Water Content ( $PWC$ )	$\text{kgm}^{-2}$
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 ( $T_s$ )	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	$\text{kgm}^{-2}$
35	81	Maximum Temperature ( $JJ$ )	K
36	82	Ship Speed ( $V_s$ )	$\text{ms}^{-1}$
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	$\text{ms}^{-1}$
50	119	Brightness Temperature ( $Tb$ )	K
51	120	Raw Radiance	K

**Table 2.4** *Continued.*

No.	Code	Name	Unit
52	121	Cloud Amount from Satellite	%
53	122	Backscatter ( $\sigma^0$ )	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
77	187	Horizontal line-of-sight wind component	$ms^{-1}$
78	174	Aerosol optical depth at 0.55 microns (AOD)	
79	163	Limb Radiances	
80	181	GEMS reactive gases, N02	
81	182	GEMS reactive gases, S02	
82	183	GEMS reactive gases, CO	
83	184	GEMS reactive gases	
84	185	GEMS reactive gases, G03	
85	175	Cloud optical depth (COD)	
86	176	Ratio of fine mode to total aerosol optical depth at 0.55 microns (RAO)	
87	177	Aerosol reflectance multi-channel (RFA)	
88	178	Aerosol optical depth multi-channel (ODA)	
89	179	Normalized Soil Moisture	0-100%
90	180	Soil Moisture	$kg^3kg^{-3}$
91	186	GHG	
92	187	GHG	
93	195	Radar doppler wind	

**Table 2.5** *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



### 2.6.2 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 2.6, 2.7 and 2.8](#). 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  $RH_{\text{err}}$  is either prescribed or modelled. More will be said about the modelled  $RH_{\text{err}}$  in [Subsection 2.6.3](#).  $RH_{\text{err}}$  is prescribed only for TEMP and SYNOP data.  $RH_{\text{err}}$  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.

### 2.6.3 Derived observation errors

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

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

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

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

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

or

$$Q_{\text{err}} = RH_{\text{err}} F_1(RH, P, T)$$

where function  $F_1$  is given by

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

Subroutine **RH2Q** is used to evaluate  $Q_{\text{err}}$ .

Surface pressure observation error  $P_{s\text{err}}$  is derived by multiplying the height observation error  $Z_{\text{err}}$  by a constant:

$$P_{s\text{err}} = 1.225 Z_{\text{err}}$$

However, the  $P_{s\text{err}}$  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  $P_t$  is adjusted for SHIP movement or not.

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

Table 2.6 Prescribed RMS observation errors for the  $u$  and  $v$  wind components ( $\text{ms}^{-1}$ ).

R RAD	SCATT		PAOB		SATEM		PILOT		TEMP		DRIBU		SATOB		AIREP		SYNOP		Obs Type	
	All	Default	All	ASCAT	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	Code
n/a	2.00	1.50	n/a	n/a	n/a	1.80	1.80	1.80	1.80	1.80	1.80	1.80	2.00	2.86	2.46	3.00	1000	Levels	(hPa)	
n/a	2.00	1.50	n/a	n/a	n/a	1.80	1.80	1.80	1.80	1.80	1.80	1.80	2.00	2.91	2.51	3.00	850			
n/a	2.00	1.50	n/a	n/a	n/a	1.90	1.90	1.90	1.90	1.90	1.80	1.80	2.00	2.96	2.56	3.00	700			
n/a	2.00	1.50	n/a	n/a	n/a	2.10	2.10	2.10	2.10	2.10	1.80	1.80	3.50	3.11	2.71	3.40	500			
n/a	2.00	1.50	n/a	n/a	n/a	2.50	2.50	2.50	2.50	2.50	1.80	1.80	4.30	3.21	2.81	3.60	400			
n/a	2.00	1.50	n/a	n/a	n/a	2.60	2.60	2.60	2.60	2.60	1.80	1.80	5.00	3.26	2.86	3.80	300			
n/a	2.00	1.50	n/a	n/a	n/a	2.50	2.50	2.50	2.50	2.50	1.80	1.80	5.00	3.31	2.91	3.20	250			
n/a	2.00	1.50	n/a	n/a	n/a	2.50	2.50	2.50	2.50	2.50	1.80	1.80	5.00	3.36	2.96	3.20	200			
n/a	2.00	1.50	n/a	n/a	n/a	2.40	2.40	2.40	2.40	2.40	1.80	1.80	5.00	3.31	2.91	2.40	150			
n/a	2.00	1.50	n/a	n/a	n/a	2.20	2.20	2.20	2.20	2.20	1.80	1.80	5.00	3.16	2.76	2.20	100			
n/a	2.00	1.50	n/a	n/a	n/a	2.10	2.10	2.10	2.10	2.10	1.80	1.80	5.00	3.06	2.66	2.00	70			
n/a	2.00	1.50	n/a	n/a	n/a	2.00	2.00	2.00	2.00	2.00	1.80	1.80	5.00	3.06	2.66	2.00	50			
n/a	2.00	1.50	n/a	n/a	n/a	2.10	2.10	2.10	2.10	2.10	1.80	1.80	5.00	3.26	2.86	2.00	30			
n/a	2.00	1.50	n/a	n/a	n/a	2.30	2.30	2.30	2.30	2.30	1.80	1.80	5.00	3.46	3.06	2.50	20			
n/a	2.00	1.50	n/a	n/a	n/a	3.00	3.00	3.00	3.00	3.00	1.80	1.80	5.70	3.76	3.36	3.00	10			

Table 2.7 Prescribed RMS height observation errors (*m*).

		SYNOP												Observation Type				
		AIREP				SHIP				Manual				Land		Code Type		
R	AD	PAOB	SATEM	PILOT	TEMP	DRIBU	SATOB	AI	SHIP	AI	SHIP	AI	SHIP	AI	SHIP	Land	Land	Levels (hPa)
n/a	All	24.0	n/a	All	4.3	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	5.6	1000
n/a	All	24.0	n/a	All	4.4	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	7.2	850
n/a	All	24.0	n/a	All	5.2	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	8.6	700
n/a	All	24.0	n/a	All	8.4	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	12.1	500
n/a	All	24.0	n/a	All	9.8	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	14.9	400
n/a	All	24.0	n/a	All	10.7	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	18.8	300
n/a	All	24.0	n/a	All	11.8	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	25.4	250
n/a	All	24.0	n/a	All	13.2	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	27.7	200
n/a	All	24.0	n/a	All	15.2	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	32.4	150
n/a	All	24.0	n/a	All	18.1	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	39.4	100
n/a	All	24.0	n/a	All	19.5	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	50.3	70
n/a	All	24.0	n/a	All	22.5	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	59.3	50
n/a	All	24.0	n/a	All	25.0	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	69.8	30
n/a	All	24.0	n/a	All	32.0	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	96.0	20
n/a	All	24.0	n/a	All	40.0	All	n/a	All	4.2	All	n/a	All	4.2	All	7.1	All	114.2	10

Table 2.8 Prescribed RMS temperature observations error (K).

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

**Table 2.9** *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 2.10** *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 2.11** *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

#### 2.6.4 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

$$F_{OE} = \sqrt{O_E^2 + P_E^2}$$

where  $F_{OE}$ ,  $O_E$  and  $P_E$  are final, prescribed and persistence observation errors, respectively. The subroutine used for this purpose is **FINOERR**.

## 2.7 DEFINITIONS

### 2.7.1 Observation characteristics: instrument specification and retrieval type

Where applicable, [Tables 2.9](#) to [2.15](#) describe in details how the ODB's instrument specification word is structured. Tables provided are for different observation types.

In [Table 2.16](#) the ODB's header retrieval word codes are described.

### 2.7.2 Vertical coordinate: pressure, satellite ID and level ID codes

In the ODB the vertical coordinate is expressed by various codes, and [Table 2.17](#) describes those codes.

**Table 2.12** *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 2.13** *TEMP instrument specification.*

Type	Bit Position	No. of Bits	Value – Description
Instrument Specification	0	10	Not Defined
Not Defined	10–30	21	Reserved

**Table 2.14** *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 2.15** *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

**Table 2.16** *Satellite retrieval codes.*

Retrieval Codes	Description
1	Clear
2	Partly Clear
3	Cloudy

**Table 2.17** *Vertical coordinate.*

Vertical Coordinate Codes	Description
1	Pressure (Pa)
2	Height (GPM)
3	Satellite Channel (numeric)
4	Scatterometer Channel (numeric)

**Table 2.18** *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 2.19** *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 2.20** *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

Also, the ODB pressure code word is expressed in terms of codes which are defined in [Table 2.18](#).

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 2.19](#).

### 2.7.3 ODB report status: events, flags and codes

The status of each ODB report is described in terms of being active, passive, rejected or blacklisted. For some microwave radiances, the additional flag `use_emiskf_only` is also used, see [1.4.2](#) and [3.4.3](#). The ODB report status word is packed with the 4 bits given in [Table 2.20](#).

There is one, 31 bits packed, word for each ODB report to account for various blacklist events. Details are given in [Table 2.21](#).

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.

**Table 2.21** *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 2.22** *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 – Redundant Report
5	1	1 – Missing Station Altitude
6	1	1 – Failed Quality Control
7	1	1 – Report Overcast IR

**Table 2.23** *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

The first ODB report event word is described in [Table 2.22](#).

The second ODB report event word holds an additional set of events which are now dependent on observation type. Details are given in [Tables 2.23](#) to [2.26](#).



Table 2.24 *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 2.25 *SATEM* report events.

Bit Position	No. of Bits	Value – Description
0	1	1 - Thinned Report
1–30	30	Not Used

Table 2.26 *SCAT* report events.

Bit Position	No. of Bits	Value – Description
0	1	1 - Report thinned in across-node direction
1	1	1 - Reported Wind Directions too Close
2	1	1 - Report in QuikScat outer swath
3	1	1 - Report Contaminated by Rain
4–30	29	Not Used

Table 2.27 *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

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 2.27](#).

#### 2.7.4 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 2.28](#) 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

**Table 2.28** *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 2.29** *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	1	1 – Cloud Contamination
29	1	1 – Rain Contamination
30	1	1 – Aerosol Contamination
31	1	1 – Missing or Not Sensible Emissivity Values

event word is observation type dependent. [Tables 2.29 to 2.32](#) describe the event words structures for the observation types that use them.

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 2.33](#).

Table 2.30 *SYNOP datum events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Bias Corrected Ps
1–30	30	Not Used

Table 2.31 *DRIBU datum events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Bias Corrected Ps
1–30	30	Not Used

Table 2.32 *TEMP datum events.*

Bit Position	No. of Bits	Value – Description
0	1	1 – Bias Corrected Value Used
1–30	30	Not Used

**Table 2.33** 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

**Table 2.34** RDB pressure (vertical coordinate) and datum flags.

Parameter	No. of Bits	Bit Position	Bit Position	No. of Bits	Value – Description
Pressure	15	0+	0	1	0 – No Human Monitoring Substitution 1 – Human Monitoring Substitution
			+1	1	0 – No Q/C Substitution 1 – Q/C Substitution
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

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 2.34](#). 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:

**Table 2.35** *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

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 2.35](#).



## Chapter 3

# Observation screening

### Table of contents

- 3.1 Introduction**
- 3.2 The structure of the observation screening**
  - 3.2.1 The incoming observations
  - 3.2.2 The screening run
  - 3.2.3 General rationale of the observation screening
  - 3.2.4 3D-Var versus 4D-Var screening
- 3.3 The independent observation screening decisions**
  - 3.3.1 Preliminary check of observations
  - 3.3.2 Blacklisting
  - 3.3.3 Background quality control
- 3.4 Screening of satellite radiances**
  - 3.4.1 Pre-screening
  - 3.4.2 Cloud and rain rejection
  - 3.4.3 Blacklisting decisions
- 3.5 Screening of Scatterometer data**
  - 3.5.1 Wind retrieval
  - 3.5.2 Quality control
  - 3.5.3 Rain contamination
  - 3.5.4 Bias corrections
  - 3.5.5 Blacklisting decisions
- 3.6 The dependent observation screening decisions**
  - 3.6.1 Update of the observations
  - 3.6.2 Global time–location arrays
  - 3.6.3 Vertical consistency of multilevel reports
  - 3.6.4 Removal of duplicated reports
  - 3.6.5 Redundancy check
  - 3.6.6 Thinning
  - 3.6.7 Compression of the ODB
- 3.7 A massively-parallel computing environment**
- Appendix A**
  - A.1 Bad reporting practice of SYNOP and TEMP reports
  - A.2 Revised background quality control for selected observations
  - A.3 Use of atmospheric motion winds

### 3.1 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](#)).

## 3.2 THE STRUCTURE OF THE OBSERVATION SCREENING

### 3.2.1 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 systems e.g. conventional (that also contains AMVs), ATOVS (that also contains AIRS radiances), geostationary radiances and IASI). 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. The observation minus model differences (the departures) are computed as described in Part II 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.

### 3.2.2 The screening run

The ECMWF 3D/4D-Var data assimilation system makes use of an incremental minimization scheme, as described in Part II. 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 Part II) 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 minimization.

### 3.2.3 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 radiance 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.

### 3.2.4 3D-Var 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 Part II. Also the current 3D-FGAT configuration (see part II) relies on 4D-screening (LSCRE4D = .TRUE.).

## 3.3 THE INDEPENDENT OBSERVATION SCREENING DECISIONS

### 3.3.1 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 (**REPR**), 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.

### 3.3.2 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 blacklist files 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 have recently been reporting in an excessively noisy or biased manner compared with the ECMWF background field.

### 3.3.3 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  $\sigma_b$ , 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

**Table 3.1** *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.

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 3.1](#) 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<sup>-1</sup>. 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.

There is no first-guess check for scatterometer data. It is demanded, though, that neither scatterometer nor model wind speed should exceed 35 ms<sup>-1</sup>, since that marks the range of validity for scatterometer wind inversion.

## 3.4 SCREENING OF SATELLITE RADIANCES

### 3.4.1 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.

### 3.4.2 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 that are assimilated in all-sky conditions are not subject to cloud or rain rejection. For other microwave radiances (e.g. AMSU-A) 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. 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 aspects of the analysis including bias parameter evolution in the VARBC.

There is a special case of infrared cloud contamination that does not lead to channels being rejected. In parallel to the setting of clear and cloudy flags, simplified cloud parameters (cloud top pressure and effective cloud fraction) are estimated from the infrared data (HIRS, AIRS and IASI) in routine `CLOUD_ESTIMATE`. In the case that a pixel is diagnosed as completely overcast and subject to some additional restrictions placed upon the altitude of the diagnosed cloud (e.g. that it is not within 100hPa of the surface), the rejection flags are NOT set. All channels in that pixel are then assimilated with the estimated cloud parameters passed to the forward operator and further evolved as local extensions to the control vector in the minimization. The additional use of overcast infrared radiances can be disabled by setting the logical variable `LCLDSINK` to false.

### 3.4.3 Blacklisting decisions

Like for 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 from 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 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 a new satellite for which we do not wish to actively assimilate the data, but wish to establish an accurate bias correction. Another application of the `FAIL(EXPERIMENTAL)` facility is its use for window channels used in the quality control of other data.

For some window channel microwave radiances over land, another setting of the blacklisting decision can be `FAIL(USE_EMISKF_ONLY)`. This means that the observation will not be used for the atmospheric analysis in a similar way as `FAIL(CONSTANT)` rejects the observation, but the observation is used to influence the emissivity Kalman Filter atlas, described in section [1.4.2](#).

### 3.5 SCREENING OF SCATTEROMETER DATA

The screening of scatterometer data involves the conversion of the backscatter measurements acquired by the instrument (triplets for ERS and ASCAT and quadruplets for NSCAT and QuikSCAT) into ambiguous  $u$  and  $v$  wind components that will actually be assimilated into the IFS (see [Section 1.5.4](#)). The (empirical) relation between wind and backscatter is described by a geophysical model function (GMF). Although in principle inverted wind components are provided as a level 2 product, at ECMWF the wind inversion is performed in house. In this way any drifts in backscatter levels can be corrected in a direct manner.

Data from the AMI instrument on ERS-2 have been used from June 1996 (with an interruption from January 2001 till March 2004), data from the SeaWinds instrument on-board QuikSCAT was used from January 2002 until November 2009 (when QuikSCAT failed), and data from ASCAT on MetOp have been assimilated from June 2007 onwards. Data from NSCAT have never been used in an operational setup, although offline assimilation experiments have been performed. From November 2010 onwards scatterometer data is assimilated as equivalent-neutral 10-metre wind, rather than (real) 10-metre wind, since the former model winds are closer related to scatterometer observations.

#### 3.5.1 Wind retrieval

Since geometry and measurement principle of ERS and ASCAT are alike, data from these instruments is processed in a similar way. The procedure for wind inversion closely follows the wind retrieval and ambiguity removal scheme originally developed for the ERS-1 scatterometer ([Stoffelen and Anderson, 1997](#)), though the original geophysical model function CMOD4 has been replaced by CMOD5 ([Hersbach et al., 2007](#)) in March 2004, by CMOD5.4 in June 2007, and by CMOD5.n ([Hersbach, 2010b](#)) in November 2010, after which scatterometer winds are assimilated as equivalent-neutral wind ([Hersbach, 2010a](#)).

For QuikSCAT the task of wind inversion is performed in the pre-screening ([PRESCAT](#)). Data are like ERS and ASCAT, provided at a resolution of 25 km. Rather than data thinning (see [Subsection 3.6.6](#)), for QuikSCAT a 50 km product is created which contains information about backscatter from the four underlying original sub-cells. 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. It is the re-sampled 50 km product that is stored in ODB. Original backscatter observations at 25 km are not available within the assimilation.

In general, the wind retrieval is performed by minimizing the distance between observed backscatter values  $\sigma_{oi}^0$  and modelled backscatter values  $\sigma_{mi}^0$  given by

$$D(u) = \sum_i^n \frac{[(\sigma_{oi}^0)^p - \sigma_{mi}^0(u)^p]^2}{k_p \left[ \sum_j^n \sigma_{mj}^0(u)^p \right]^2} \quad (3.1)$$

For ERS and ASCAT 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 and ASCAT 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  $\sigma_m^0$  is for ERS and ASCAT data based on the CMOD5.4 model function. For NSCAT data the NSCAT-2 GMF has been utilized. For QuikSCAT data, the choice of GMF is handled by a logical switch LQTABLE. 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 and ASCAT data, steps of  $0.5 \text{ ms}^{-1}$ ,  $5^\circ$  and  $1^\circ$ , respectively. For NSCAT and QuikSCAT data the corresponding values are  $0.2 \text{ ms}^{-1}$ ,  $2.5^\circ$  and  $1^\circ$ . ERS and ASCAT use the same table, which is read in the initialisation subroutine [INIERSCA](#). For QuikSCAT, inversion takes place in the [QSCAT25TO50KM](#) program in the [PRESCAT](#) task.

### 3.5.2 Quality control

The wind inversion involves some quality control. For ERS (**ERS1IF**),  $k_p$  must for each antenna be below 10%, and a missing packet number must be less than 10 to ensure that enough individual backscatter measurements have been averaged for estimating the value.

For ASCAT (**ASCATIF**)  $a$  in the product provided land fraction must be zero for each backscatter measurement. No restriction on  $k_p$  is imposed, other than that values should be non missing. It is checked whether two other provided quality flags (‘sigma0 usability’ and ‘kp quality’) have acceptable values. However, no quality control decisions are made on these two indicators for the moment, since so far, they have not been fully calibrated and validated by EUMETSAT.

For QuikSCAT, from 38 across-track 50 km cells, the outer 4 at either side of the swath are, due to their known reduced quality rejected. 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  $135^\circ$  (routine **SCAQC**).

After wind inversion, a further check is done on the backscatter residual associated to the rank-1 solution (also called ‘distance to the cone’). This misfit 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, a triplet is rejected when the cone distance exceeds a threshold of three times its expected value. For QuikSCAT and ASCAT data such a test is not performed.

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 and ASCAT, cell-wise averages are calculated for the default 4D-Var observation time slot (30 minutes) 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 evaluated in the pre-screening (**DCONE\_OC**), using a threshold of 1.45 for any of cells between 5 and 34.

### 3.5.3 Rain contamination

Thanks to the usage of C-band frequency, rain contamination is mild for ERS and ASCAT. For QuikSCAT and NSCAT, which operate in Ku band, rain contamination is a serious issue.

For QuikSCAT the check on rain contamination occurs in the pre-screening and is imposed on the original 25 km observations. Any 25 km rejected cell is not used in the determination of the 50 km wind product. When more than one 25 km sub-cell is rejected, the entire 50 km product is rejected (decision made in **SCAQC**).

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  $2 \text{ m}^2 \text{ hr}^{-1}$ ; 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 (more aggressive) 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. It is given by

$$L_{\text{rain}} = \{ \text{nof\_rain\_index} + 200 \text{ mp\_rain\_probability} > 30 \}.$$

Both *mp\_rain\_probability* and *nof\_rain\_index* are provided in the original 25 km 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.



### 3.5.4 Bias corrections

For ASCAT and ERS, bias corrections are applied, both in terms of backscatter (before wind inversion) and wind speed (after inversion), particularly to compensate for any change in the instrumental calibration and to ensure consistency between the retrieved and model winds. The backscatter and wind-speed bias corrections are defined by dedicated files read in the initialization subroutine **INIERSCA**. Files are in principle model-cycle and date dependent. Currently for ERS-2, the appropriate files have no effect (i.e. containing only unity correction factors and zeros), since the CMOD5.4 GMF was tuned on ERS-2. For ASCAT, though, the usage of bias corrections is essential, since the backscatter product for this instrument has been calibrated differently from ERS. The bias correction file for backscatter has been updated every time a change in the calibration of ASCAT was imposed by EUMETSAT.

For QuikSCAT data no bias corrections in  $\sigma^0$  space is applied, though, wind-bias corrections are made. This also takes place in the pre-screening. Corrections are performed in three steps. First of all, wind speeds are slightly reduced according to:

$$v' = 0.2 + 0.96 v.$$

Where  $v$  is the wind speed as obtained from inversion (3.1) The addition of  $0.2 \text{ ms}^{-1}$  is used in the operational configuration, where scatterometer data is assimilated as equivalent neutral wind. In case this is not desired (expressed by `LSCATT_NEUTRAL=false`) only the rescaling factor of 0.96 is used. It was observed that the residual bias between QuikSCAT winds and ECMWF first guess winds depends on the value of `mp_rain_probability`. 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. over rain-free sub-cells). The maximum allowed correction is  $2.5 \text{ ms}^{-1}$ , 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  $19 \text{ ms}^{-1}$  the following correction is applied:

$$v''' = v'' - 0.2(v'' - 19.0).$$

### 3.5.5 Blacklisting decisions

In order to screen on sea-ice contamination, scatterometer data are removed (within the blacklist mechanism) whenever the model sea-ice fraction exceeds 1% or the model sea-surface-temperature analysis is below 273.15 K. Land is removed by imposing that the model land-sea mask should not exceed 10%.

## 3.6 THE DEPENDENT OBSERVATION SCREENING DECISIONS

### 3.6.1 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.

### 3.6.2 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).

### 3.6.3 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.

### 3.6.4 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.

### 3.6.5 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**).

### 3.6.6 Thinning

Finally, a horizontal thinning is performed for the **AIREP**, radiances (**ATOVS**, **AIRS**, **IASI**), **GEOS**, **SATOB**, **ERS** and **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 radiances, **GRAD**, **SATOB**, **ERS** and **ASCAT** reports are each done in two stages controlled by **THINN**. For radiances (**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 radiances 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 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.

For ERS and ASCAT scatterometer data, the above described thinning algorithm is only applied along track. In across-track direction, backscatter data from these platforms are provided into wind-vector cells (WVC) with a spatial resolution of 25 km. In this direction, data is thinned by selecting predefined wind-vector cells (subroutine **SCAQC**). For ERS, from 19 cells, only 3, 7, 11, 15 and 19 are regarded (cells 1 and 2 are of known lower quality). For ASCAT, from 42 cells (two swaths of 21 cells each) only cells 1, 5, 9, 13, 17, 21, 22, 26, 30, 34, 38 and 42 are used. After this across-track thinning, the generic thinning algorithm is applied to the remaining cells in along-track direction. QuikSCAT data (also provided on a 25 km grid) are not thinned. Instead, a 50 km wind product is determined from backscatter data from four underlying 25 km cells, each given a reduced weight of one fourth (see [Subsection 3.5](#)).

Scatterometer winds are besides thinning subject to a high-wind rejection test with an upper-wind speed limit set to  $35 \text{ ms}^{-1}$  to both the scatterometer and background winds (**FGWND**).

### 3.6.7 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 radiance data, since after the screening there are only 10–20% of the radiance 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.

## 3.7 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 3.1](#). The special cases are the following ones.

- AIREP wind observations with zero wind speed are rejected if the background wind exceeds  $5 \text{ m s}^{-1}$ .
- 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  $55^\circ$  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  $20^\circ\text{N}$ .
- For Meteosat ( $0^\circ$  mission) instead of  $20^\circ\text{N}$  this threshold is  $35^\circ\text{N}$  to allow usage of AMW over north Africa.
- For Meteost ( $63^\circ$  mission) the use of AMW has been extended over Asia if above 500 hPa. This is restricted for longitudes east of  $30^\circ\text{E}$ .
- AMW are blacklisted over the Himalayas as a precautionary measure.
- AMW over land south of  $20^\circ\text{N}$  ( $35^\circ\text{N}$  for Meteosat) is used if above 500 hPa.

Check on satellite ( $35^\circ\text{N}$  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  $ERRLIM_j * ZREJMOD$  of its estimated variance, as given by the following expression:

if  $[D^2 > (sfg^2 + sobs^2) * ERRLIM_j * ZREJMOD]$  then flag =  $j$  where  $D^2 = 1/2 (Du^2 + Dv^2)$  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,  $2 \text{ m s}^{-1}$  for levels below 700 hPa included,  $3.5 \text{ m s}^{-1}$  at 500 hPa,  $4.3 \text{ m s}^{-1}$  at 400 hPa and  $5 \text{ m s}^{-1}$  for all levels above;  $ERRLIM_j$  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  $4 \text{ m s}^{-1}$  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 =  $0.075 - 0.00125 * \text{SPDfg}$  all others
- ZREJMOD = 0.0 if SPDfg >  $60 \text{ m s}^{-1}$  (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  $j = 3$ .



## 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., 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.
- Bauer, P., Geer, A. J., Lopez, P. and Salmond, D. (2010). Direct 4D-Var assimilation of all-sky radiances: Part I. Implementation. *Q. J. R. Meteorol. Soc.*, **136**, 1868–1885.
- 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. (2014). Reformulation of the emissivity kalman filter atlas. *RD Internal Memorandum*, p. 14pp, available online: <http://intra.ecmwf.int/publications/library/do/references/show?id=1471>.
- 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.
- Buck, A. L. (1981). New equations for computing vapor pressure and enhancement factor. *J. Appl. Meteorol.*, **20**, 1527–1532.
- 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.
- 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.
- Deblonde, G. and English, S. (2001). Evaluation of the fastem-2 fast microwave oceanic surface emissivity model. In *Tech. Proc. ITSC-XI*, pp. 67–78, Budapest.

- Engelen, R. J., Andersson, E., Chevallier, F., Hollingsworth, A., Matricardi, M., McNally, A., Thépaut, J.-N. and Watts, P. (2004). Estimating atmospheric CO<sub>2</sub> 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*.
- Geer, A. J., Baordo, F., Bormann, N. and English, S. (2014). All-sky assimilation of microwave humidity sounders. *ECMWF technical memorandum, in preparation*.
- Geer, A. J. and Bauer, P. (2011). Observation errors in all-sky data assimilation. *Q. J. R. Meteorol. Soc.*, DOI: 10.1002/qj.830.
- Geer, A. J. and Bauer, P. (2012). Assimilating AMSU-A radiances sensitive to cloud and temperature using the all-sky approach. *ECMWF technical memorandum 670*.
- Geer, A. J., Bauer, P. and Lopez, P. (2008). Lessons learnt from the operational 1D+4D-Var assimilation of rain- and cloud-affected SSM/I observations at ECMWF. *Q. J. R. Meteorol. Soc.*, **134**, 1513–1525.
- Geer, A. J., Bauer, P. and Lopez, P. (2010). Direct 4D-Var assimilation of all-sky radiances: Part II. Assessment. *Q. J. R. Meteorol. Soc.*, **136**, 1886–1905.
- Geleyn, J.-F. (1988). Interpolation of wind, temperature and humidity values from the model levels to the height of measurement. *Tellus*, **40**, 347–351.
- 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. (2010a). Assimilation of scatterometer data as equivalent-neutral wind. *ECMWF Tech. Memo. No. 629*.
- Hersbach, H. (2010b). Comparison of c-band scatterometer cmod5.n equivalent neutral winds with ecmwf. *J. Atmos. Oceanic Technol.*, **27**, 721–736.
- Hersbach, H. (2011). Sea surface roughness and drag coefficient as functions of neutral wind speed. *J. Phys. Oceanogr.*, **41**, 247–251.
- Hersbach, H., Stoffelen, A. and de, S. H. (2007). An improved C-band scatterometer ocean geophysical model function: CMOD5. *J. Geophys. Res.*, **112**, doi:10.1029/2006JC003743.
- Hocking, J. (2014). Interpolation methods in the rttov radiative transfer model. *Forecast Research Technical Report No. 590*.
- 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/>.
- 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.
- 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*.
- Karbou, F., Gérard, E. and Rabier, F. (2006). Microwave land emissivity and skin temperature for AMSU-A and -B assimilation over land. *Q. J. R. Meteorol. Soc.*, **132**, 2333–2355.

- Kelly, G. (2007). The evaluation of the hirs/4 instrument on metop 2 by using the ecmwf data assimilation system to provide a reference to compare with other sensors. *ECMWF Tech. Memo. No. 523*.
- 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*.
- Krzeminski, B., Bormann, N., Karbou, F. and Bauer, P. (2009). Improved use of surface-sensitive microwave radiances at ECMWF. In *Proceedings of 2009 EUMETSAT Meteorological Satellite Conference*, Bath, United Kingdom.
- 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.
- 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, Reading, 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, Reading, 6–9 November 1984.
- Lönnberg, P. and Shaw, D. (Eds) (1987). *ECMWF Data Assimilation Scientific Documentation, Research Manual 1*.
- Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194.
- 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.
- Lupu, C. and McNally, A. (2012). Assimilation of cloud affected radiances from meteosat-9 at ecmwf. *ECMWF/EUMETSAT fellowship reports No. 25*.
- 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. (2009). The direct assimilation of cloud affected satellite infrared radiances in the ecmwf 4d-var. *Q. J. R. Meteorol. Soc.*, **135**, 1214–1229.
- 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 No. 83*, pp. 2–7.
- McNally, A. P. and Watts, P. D. (2003). A cloud detection algorithm for high spectral resolution infrared sounders. *Q. J. R. Meteorol. Soc.*, **129**, 3411–3423.
- 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.
- 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.

- 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.
- Stoffelen, A. and Anderson, D. (1997). Ambiguity removal and assimilation of scatterometer data. *Q. J. R. Meteorol. Soc.*, **123**, 491–518.
- Tomassini, M., LeMeur, D. and Saunders, R. W. (1998). Near-surface satellite wind observations of hurricanes and their impact on ECMWF model analyses and forecasts. *Mon. Wea. Rev.*, **126**, 1274–1286.
- 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.