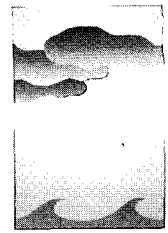# A Simulated Future Atmospheric Observation Database Including ATOVS, ASCAT, and DWL

Bernd Dieter Becker, Hervé Roquet,* and Ad Stoffelen+

European Centre for Medium-Range Weather Forecasts, Reading, Berkshire, United Kingdom

ABSTRACT

A database for study of the impact of Doppler wind lidar data on numerical weather prediction in Observation System Simulation Experiments was created. Five Doppler wind lidar scenarios, TIROS Operational Vertical Sounder, Advanced TIROS Operational Vertical Sounder, Advanced Scatterometer, and all conventional observation types with a realistic distribution in time and space have been successfully simulated. A 30-day run of the ECMWF forecast model was used as a physically sound reference state. This "true" atmospheric state was sampled at the observation positions and times. The simulated true variables were mapped onto the "measured" variables, and a mix of random and gross errors with realistic statistical characteristics was added. The simulated observations were validated by comparison with existing data where available.

1. Introduction

The accuracy of a forecast from a numerical weather prediction (NWP) system depends on a very accurate description of the initial state of the atmosphere and a realistic computer model to simulate the subsequent atmospheric evolution. The main variables that need to be observed in order to provide an adequate description of the atmosphere on scales represented by most of the present global NWP models are the three-dimensional fields of temperature, humidity, and wind at a 50–100-km resolution in the horizontal and 1 km in the vertical from the surface to about 30-km altitude. Space-based observing systems offer realistic hope of meeting the requirements of operational meteorology for a global coverage of observations with adequate spatial and temporal resolutions. Besides the observation coverage, spatial and

*Current affiliation: Météo-France, Toulouse, France.
+Current affiliation: KNMI, De Bilt, the Netherlands.
Corresponding author address: Bernd Dieter Becker, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire, RG2 9AX, United Kingdom.
E-mail: BBecker@ecmwf.int
In final form 31 May 1996.

temporal resolutions, the accuracy of the analysis of the initial state is highly dependent on the observation accuracy. Over the oceans, which cover most of the earth's surface, the horizontal and vertical coverage of "conventional" observations (i.e., from surface stations/platforms, balloons, and aircraft) is sparse. In the mid- and high latitudes, the temperature and wind fields are coupled through the dynamical equations governing atmospheric motion. However, there is a significant component of the flow on small horizontal scales that is not coupled, making measurements of wind useful and important in the presence of temperature observations. In the Tropics, the coupling is much weaker and direct observations of wind are even more important.

The current wind observation coverage is far from the above requirement. One of the promising future instruments is a spaceborne Doppler wind lidar (DWL). It measures the Doppler shift of a refracted signal due to moving aerosol and cloud particles along a line of sight (LOS). A vertical wind component profile can be inferred. By probing in different azimuthal directions, the full horizontal wind vector field may be resolved. Such an instrument could profile the atmosphere from 20-km height to the surface with a high vertical resolution. A comprehensive introduction to lidar-measured wind from space is given by

Baker et al. (1995) and Betout et al. (1989). However, the costs of a high resolution and accurate DWL system are considerable and cost/benefit analyses need to be made. A frequently used technique to evaluate the impact of observing systems on NWP is to perform Observing System Simulation Experiments (OSSEs). In such experiments, two data assimilation runs are performed in parallel, one including and the other excluding the observing system under investigation. The parallel sets of analyses and subsequent forecasts are then compared to assess the impact of the observing system under investigation. The European Space Agency (ESA) has initiated two study contracts to prepare a database to assess the impact of DWL data in OSSEs that were recently completed by Stoffelen et al. (1994) and Roquet et al. (1995). This article provides an overview of what was done and introduces the final product to potential users.

First, we mention some opinions and recommendations on the prerequisites for OSSEs from the literature. OSSEs usually (Andersson et al. 1991) provide the clearest answers in the following circumstances:

- the simulated data have similar errors as the real data used in the operational analyses;
- significant meteorological activity over the area and period of the study is present; and
- the data under investigation are the only data available in the area of interest—that there is no data redundancy [other than necessary for quality control (QC)].

The main drawback of the OSSE approach is the uncertainty in describing the error structure of a yet-to-be-deployed observing system and the use of numerical models to provide both the "observed" atmosphere and the verification "truth." The latter is known to be the identical or fraternal twin problem, which results from the fact that even two different models may be more like each other than like the atmosphere (Carr et al. 1993). Therefore, we did not attempt to run OSSEs with the identical system but concentrated on the simulation and validation of the database. The assumed error structure is described later in this paper.

TABLE 1a. The OSSE database contains these simulated observations over a 30-day period at each observation point. The abbreviations are pp for pressure, dd and ff for wind direction and speed, T for temperature, rH for relative humidity, T_{dew} for dewpoint temperature, T_{bright} for brightness temperature, Z for geopotential, HLOS for horizontal line-of-sight wind, and σ_0 for radar backscatter measurement.

Observation	pp	dd	ff	T	rH	T_{dew}	T_{bright}	Z	HLOS	σ_0	Level
SYNOP	×	×	×	×	×	×					Surface
SHIP	×	×	×	×	×	×					Surface
AIREP	×	×	×	×							Flight level
SATOB	×	×	×	×							Cloud top
DRIBU	×	×	×	×	×	×					Surface
TEMP	×	×	×	×		×		×			31 levels
PILOT	×	×	×								31 levels
TOVS							×				25 channels
PAOB	×										Surface
DWL									×		0–20 km
ATOVS							×				38 channels
ASCAT										×	Sea surface

The observations should be simulated at the observation site and include as realistic an error structure as possible. Both random and systematic errors, including where appropriate the instrument's response to scales of motion not resolved by current analysis–forecast systems, should be included, as well as simulations of any other future observing systems (Arnold and Dey 1986).

Without the actual instrument being operational, data must be simulated to represent what the observing system would sample. The database must be created to represent the ability of the instrument to take measurements that can be assimilated into analyses for global forecast models (Rohaly and Krishnamurti 1993).

Controversy in the interpretation of OSSEs arises because assimilation systems do not always exploit observational data to the best possible effect toward an improved forecast. In data assimilation systems, assumptions on model and observation error characteristics have to be made in order to make an analysis. (Hollingsworth and Lönnberg 1989). The provision of both the truth and the time- and space-dependent

observation errors in this database allows the testing of assimilation systems.

The original purpose of this database was to assess potential impact of DWL observations on NWP. The data are available in their full natural coverage in space and in time. For completeness, we have included simulated observations from additional satellite systems, such as Advanced Scatterometer (ASCAT) under development by ESA and Advanced Tiros Operational Vertical Sounder (ATOVS) on the next generation of National Oceanic and Atmospheric Administration (NOAA) polar orbiters that might be used to supplement DWL data. As such, this database is the most extensive ever produced. Furthermore, special care was taken to generate realistic observation error characteristics for all data. The main innovation here was to include gross errors. Gross errors are caused by instrument failure or report transmission errors, but also by the fact that a measurement is not representative of the NWP model resolution. When a gross error occurs, the observation does not relate to the atmospheric state and is

TABLE 1b. The OSSE database contains these nature run values over a 30-day period at each observation point. The abbreviations are pp for pressure, u and v for wind vector components, T for temperature, rH for relative humidity, q for specific humidity, HLOS for horizontal line-of-sight wind, TCC for total cloud cover, T_{bright} for brightness temperature, CLW for cloud water content, CC for fractional cloud cover per grid volume, and σ_0 for radar backscatter measurement.

Observation	pp	u	v	T	rH	TCC	T_{bright}	CLW	CC	σ_0	Level
SYNOP	×	×	×	×	×						Surface
SHIP	×	×	×	×	×						Surface
AIREP	×	×	×	×							Flight level
SATOB	×	×	×	×	×						Cloud top
DRIBU	×	×	×	×	×						Surface
TEMP	×	×	×	×							31 levels
PILOT	×	×	×								31 levels
TOVS						×	×				25 channels
PAOB	×										Surface
DWL		HLOS			q				×	×		0–20 km
ATOVS						×	×				38 channels
ASCAT										×	Sea surface

therefore potentially damaging for data assimilation. Forecast skill is known to be sensitive to gross error elimination procedures, that is, quality control in critical atmospheric conditions (Jarraud et al. 1989). We simulated gross error levels for all data types based on the experience with operational QC procedures. In previous databases prepared for OSSEs, no account was taken of gross error rates. Furthermore, this database is an improvement on older OSSE databases, as it takes account of the full three-dimensional cloud information at every model gridpoint for the simulation of DWL, TIROS Operational Vertical Sounder (TOVS), and ATOVS (Rohaly and Krishnamurti 1993).

This paper is only concerned with the simulation of the observations, and therefore OSSEs using this database are not described.

A description of how the OSSE database was created is given in section 2. [For more technical detail refer to Becker and Roquet (1995).] In section 3, a brief description of validation tasks and some of the results will be presented. Contents of the simulated observation database are described in section 4. Access to the OSSE database is granted by ESA on request (explained in the appendix).

2. How the OSSE database was created

A suitable candidate for the creation of an OSSE database under the above-mentioned constraints is the high-resolution ECMWF Integrated Forecasting System (IFS) (Courtier et al. 1991; Ritchie et al. 1995), which for data assimilation purposes is capable of running a forecast and comparing it with observations at the same time.

Using the IFS, the following observing systems were simulated for a period of 30 days from 5 February to 6 March 1993:

- Conventional observations
  - surface observations (SYNOP)
  - ship observations (SHIP)
  - observations from drifting and moored buoys (DRIBU)
  - radiosonde observations (TEMP)
  - pilot balloon observations (PILOT)
  - Southern Hemispheric surface-pressure observations derived from imagery and ancillery information (PAOB)
  - aircraft observations (AIREP)

- Present satellite systems
  - TIROS Operational Vertical Sounder (TOVS)
  - satellite cloud-tracked winds (SATOB)
- Future satellite systems
  - Doppler wind lidar (DWL)
  - Advanced SCATterometer (ASCAT) ·
  - Advanced TOVS (ATOVS).

The spatial coverage is comparable to a normal day's data coverage from the global observation network available via the Global Telecommunication System (GTS). The employed error characteristics were derived from many years experience from observation QC procedures at the U.K. Meteorological Office (UKMO) and at ECMWF (Lorenc et al. 1991; Gandin et al. 1993; Eyre 1990; Eyre et al. 1993; J. Eyre 1994, personal communication; Ingleby and Parrett 1994; R. Graham 1994, personal communication; ECMWF 1992). Statistics on observation accuracy and reliability of both centers were compared in detail and generally found in agreement. They include random observation errors, and where appropriate representativeness errors, to take into account the instrument's response to scales of motion not resolved by current analysis–forecast systems (Stoffelen et al. 1994; Roquet et al. 1995). A summary of the actual measurements simulated is given in Tables 1a and 1b. Observation coverage plots can be viewed in Stoffelen et al. (1994) and Roquet et al. (1995).

The ECMWF IFS was run at full resolution to provide information on the atmospheric state assumed to be without error at observation positions. This is called the nature run. In a postprocessing step, an error model converts these nature run values to true observations by applying gross and Gaussian errors as described briefly in section 2e and in more detail in Stoffelen et al. (1994) and Roquet et al. (1995).

The choice of using a numerical atmospheric model to conduct this task is based on the fact that it provides a consistent evolution in space and time for parameters that are not analyzed by a data assimilation system (i.e., cloud information). In most current assimilation systems, an analysis is available only every 6 h, which would leave us with considerable difficulties to simulate contiguous observations with a temporal resolution of the order of seconds. The IFS meets these requirements. It even provides an integral framework for these tasks. We decided to use the full original data coverage from conventional observations but employ orbit simulators for the DWL, ASCAT, ATOVS, and TOVS instruments for the whole period.

We could have used the existing TOVS data coverage from National Environmental Satellite, Data and Information Service (NESDIS), but to be independent from technical problems in the TOVS data reception over this rather long period in time and to be consistent with the treatment of the other orbiting satellite platforms, we chose the orbit simulator. The tasks listed below and depicted in the data flow diagram of Fig. 1 are described in more detail below:

• an orbit simulator for conical and crosstrack scanning satellites,
• an observation preprocessor to set up observation files,
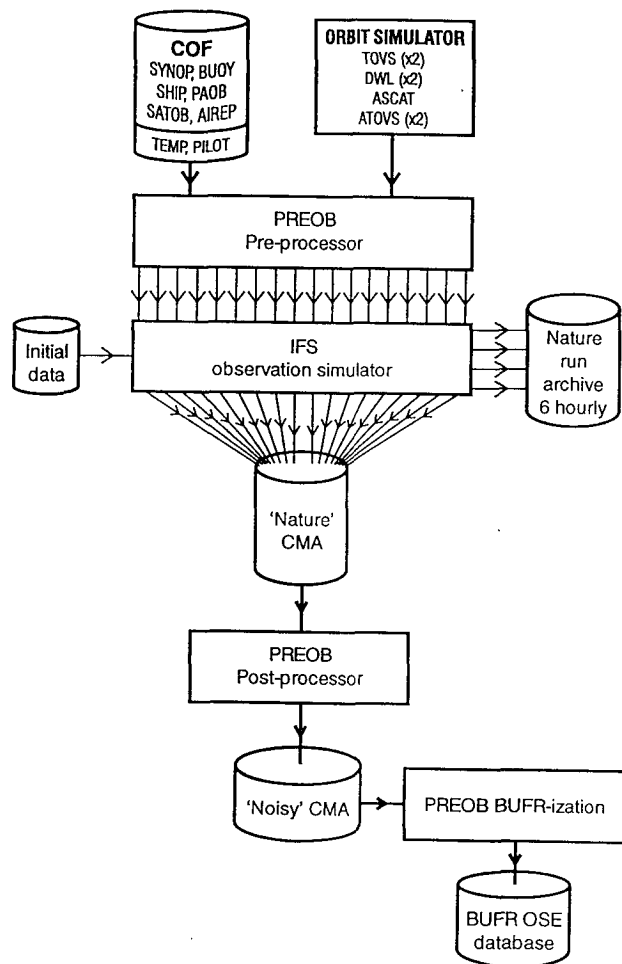• a forecasting system in observation comparison mode to provide true observations, and

FIG. 1. Schematic of the observation simulation system. COF: Comprehensive Observation File, contains all available observations. IFS: Integrated Forecasting System; CMA: Central Memory Array, medium for communication between model and observations; PREOB: creates CMA from COF and from the orbit simulator output and manipulates observations.

• an observation postprocessor to
  - apply a neutral (zero error) postprocessor
  - apply an error model to the true observations
  - convert observations into Binary universal Format Representation (BuFR).

a. The orbit simulator

This section describes the spatial distribution of the simulated satellite data. The Laboratoire de Météorologie Dynamique (LMD) provided a program to calculate the positions of fields of view (FOVs) of a conical scanning instrument on a satellite platform in polar orbit with given inclination and altitude for a prescribed period in time (P. Flamant 1994, personal communication). To provide a uniform ground coverage for the DWL instrument (Fig. 2), we simulated the conical scanning instrument with an odd number of shots (19) per two mirror cycles (20 s). For TOVS and ATOVS, we simulated the 120-km resolution TOVS NESDIS product. That is, 18 FOVs across track and one in every third scanline that is randomly thinned by 50%. The ASCAT instrument is based on a design for the payload of the European meteorological polar satellites (METOP). Each FOV is simulated as seen under its individual (node dependent) incidence angle on either side of the satellite track. FOVs over land or ice are not simulated.

b. The observation preprocessor

This tool sets up the observation files that ultimately communicate with the forecast model. Because of the large volumes of simulated data, there were con-
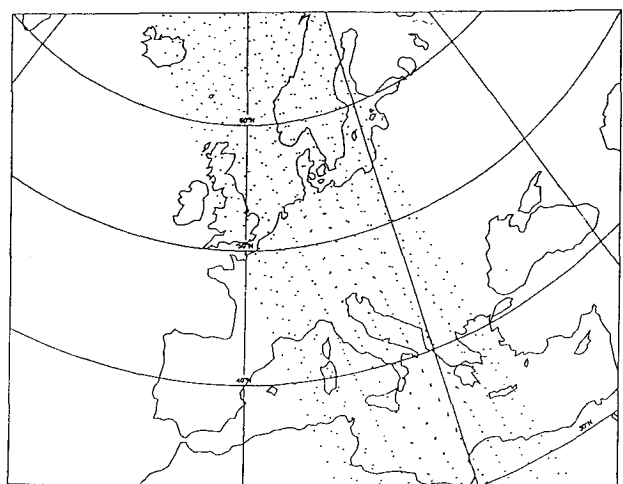

FIG. 2. DWL conical scan pattern with 19 shots per double mirror cycle, representing a shot frequency of 0.95 Hz. Each dot represents one profile.

siderable technical problems to surmount. Figure 3 shows the observation data count per observation type for a typical 24-h period. The original actual conventional measurement values are overwritten with 0 and merged with the orbit simulator output, to provide only characteristic data coverage by time and location in latitude, longitude, and pressure (channel for TOVS and ATOVS). This set of files is then compared with the simulated truth from the forecast fields in the forecasting system at every hour, to account for the asynoptic nature of observations, and the simulated true value is stored. This value will be called the nature run value.

c. The nature run truth

The ECMWF IFS global spectral weather forecasting model cycle 12r1, using triangular truncation at wavenumber 213, 31 levels in the vertical, and envelope orography, was run from an initialized analysis in normal forecast mode over the whole 30-day, 720-h period. This resolution is equivalent to a 62.5-km horizontal gridpoint spacing (Hortal and Simmons 1991). The smallest resolved half-wavelength is 94 km. The 31 vertical model levels are defined by hybrid coordinates (Simmons and Strüfing 1981) with between three and eight levels in the boundary layer at approximately 33, 150, 360, 640, 970, 1360, 1800, and 2290 m above the surface. The physics package

consists of the radiation scheme by Morcrette (1990), the mass-flux convection scheme by Tiedtke (1989), and the gravity wave drag scheme by Miller et al. (1989). The boundary layer scheme and the land surface scheme are described by Viterbo and Beljaars (1995). The TOVS and ATOVS brightness temperatures are calculated using the radiative transfer model within the IFS (Eyre 1991) and assumes a surface emissivity of unity for all channels. The scatterometer backscatter measurements were calculated with CMOD4 (Stoffelen and Anderson 1995).

The nature run truth is the forecast model output. The nature run fields are postulated to be the true atmospheric state.

d. The postprocessor

The postprocessor is run to calculate statistics, apply an active or a neutral (zero error) postprocessing, and includes the step to convert the observations into BuFR. For TOVS and ATOVS data, the postprocessor includes the decision-making logic on the cloud-clearing route, on which different error characteristics are applied in the error model (Stoffelen et al. 1994).

For DWL data, we simulated both detection error, which is dependent on instrument and atmospheric optical properties, and representativeness error, which is a measure of how the spatial and temporal scales of a particular observation represent the spatial and temporal scales resolved by the equivalent simulated model variable. Detection error was derived from a universal profile of clear-sky backscatter and transmission computed from the nature run relative humidity, cloud cover, cloud water/ice content, and a climatological temperature profile that provides a signal-to-noise ratio (SNR) that is linked to the LOS wind error by the so-called Zrnic equation (Courtier et al. 1992). The detection error and representativeness error have been squared and added, and the square root has been taken to give the final standard deviation of error for the horizontal line-of-sight wind (HLOS). [See Stoffelen et al. (1994) for more details.]
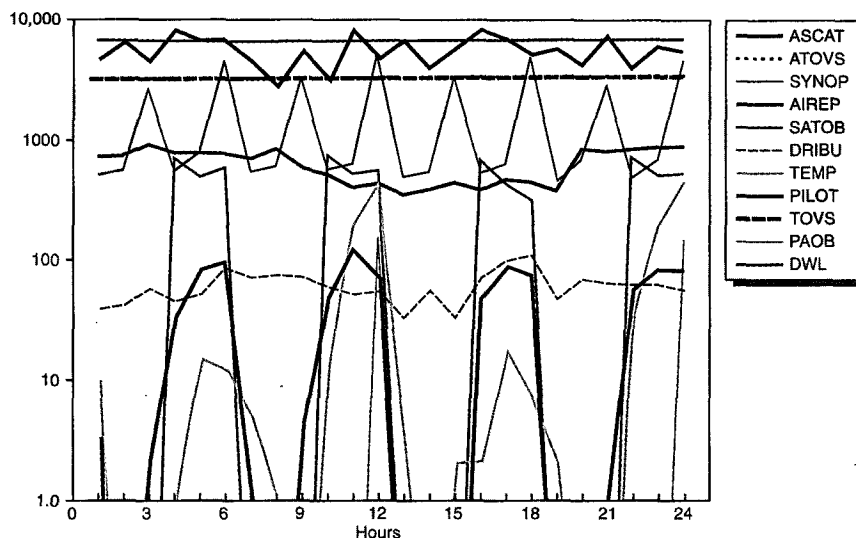

Fig. 3. Typical observation coverage in time for a 24-h period. Mind the logarithmic scale for the observation count on the left. The ASCAT observation count reflects the varying subsatellite track over land/ice and open sea, whereas the TOVS, ATOVS, and DWL observation counts reflect a uniform coverage along the swath. The other observations are taken from the GTS and show the same counts as were reported at the time (25 February 1993).

ASCAT data are simulated from the 10-m nature run wind after application of a Gaussian error with a standard deviation of 1.6 m s^{-1}, which is then transposed into σ_0 space using the transfer function CMOD4 (Stoffelen and Anderson 1995) where a further wind-speed-dependent error and a constant representativeness error are applied (Roquet et al. 1995).

For the other observation types, the statistics applied in the error model are deduced from experience in data assimilation at the UKMO and at ECMWF and are deduced from observation QC procedures (Lorenc et al. 1991; Gandin et al. 1993; Ingleby and Parrett 1994; R. Graham 1994, personal communication; ECMWF 1992). Neutral postprocessing is a mere change of the position of the nature run value in the observation file to allow for an inner consistency check of the nature run production. Active postprocessing is the application of the error model to all observations.

The data volume reduction was a main effort. Converting the data into BuFR allows an efficient way to store and retrieve huge data files. We achieved high compression rates of 1/40 and better, depending on data type.

e. The error model

The error ε applied to each datum is determined by a standard deviation σ, a probability of gross error Pg, and a gross error range Rg. Depending on a random number, in a fraction 1-Pg of cases a random Gaussian error with the standard deviation σ or alternatively a random error uniform in the range Rg is added to the nature run value:

$$\varepsilon = \begin{cases} G\,\sigma\,; \rho \le \mathrm{Pg} \\ P\,Rg\,; \text{otherwise}\,. \end{cases}$$

Here G is a function providing a random number from a Gaussian distribution with a mean of zero and a standard deviation of one; P is a function providing a random number from a homogeneous distribution in the range $(-0.5, 0.5)$; ρ is similar, but operates in the range $(0, 1)$. The total error distribution is shown in Fig. 4. Tables 2 and 3 show σ, Pg, and Rg for most of the observation types. See also Stoffelen et al. (1994) and Roquet et al. (1995) for more detail.

3. Validation tasks

After the application of the error model the simulation is complete. However, the task of validating the

simulation procedure and the simulated values was at least as demanding as the production itself. Most of the validation tasks were run alongside the production, every 24 h for 30 days, and their results were carefully examined (Becker and Roquet 1995). Some examples of the validation of this database are briefly outlined below.

a. Orbit simulator

To prove contiguously simulated orbits, they were plotted and compared with true coverage plots if they exist.

b. Observation preprocessor

Data coverage plots validated the integrity of the simulated orbits, and comparison with operational data coverage plots gives further checks for missing data, etc. The number of measurements plotted as a time series allowed monitoring of the continuity of the observation file production.

c. Nature run truth

After a 1-day forecast followed by neutral (zero error) postprocessing, the same set of observations underwent an identical forecast comparison for an integrity check. The measure of distance between the forecast and the observed atmospheric state for all observation types was 0, showing that the nature run was consistent and reproduceable. Simple monitoring of the nature run was accomplished by plotting standard deviation and mean profiles on pressure levels (Fig. 5) for DWL wind components (m s^{-1}), relative humidity (%), cloud cover (%), and cloud liquid
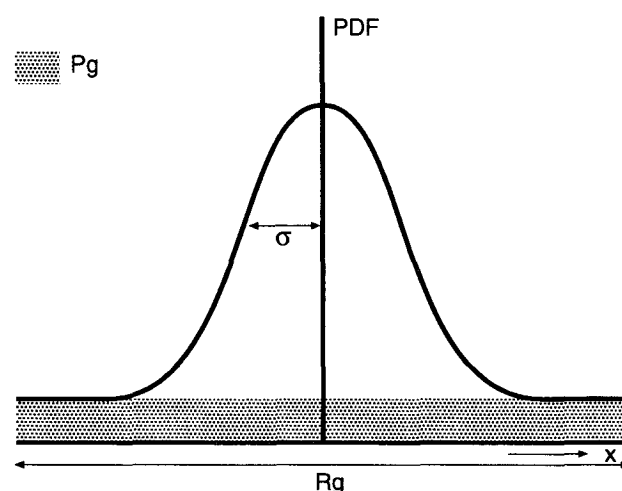

Fig. 4. Assumed error model. PDF: probability density function; σ: standard deviation; Rg: gross error range; Pg: gross error probability; x: observation domain.

water (g kg⁻¹) averaged over the Northern Hemisphere north of 20°N. The figure refers to all data in a period of one day in a DWL scenario with 800-km orbit. The numbers on the right-hand side of the individual graphs indicate the data count per level. Figure 6 shows the simulated clear column brightness temperatures by channel number for the ATOVS instrument over sea, as they are simulated from the radiative transfer model within the IFS (Eyre 1991).

d. Postprocessor

Most important here is to check the correct application of Gaussian and gross errors in the error model for all observation types. The expectation is that as long as the theoretical error standard deviation applied in the error model is constant with the atmospheric state, it is identical with the root-mean-square (rms) of the applied error. The mean of the applied error should approach zero when a higher number of samples is taken. In Fig. 7 the rms of the simulated errors is plotted against the standard deviations applied in the error model. Statistics on the Gaussian part of the error model are presented on the left-hand side of Fig. 7. There is an exact match between the theoretical and applied errors for TEMP wind u component (Fig. 7a), and the mean of the applied error is very close to zero as the theoretical standard deviation of error does not vary with atmospheric state. The expectation is met very well.

The SNR for the DWL varies with the optical properties of the penetrated column. As a consequence, the standard deviation of the theoretical error in HLOS is not constant but varies with the varying cloud and moisture fields throughout the troposphere. In Fig. 7c above 170 hPa, the match between the theoretical error (thin solid line) and the applied error (thick dashed line) is good, as less cloud particle backscatter here lowers the SNR (the amount of aerosol particles is constant at each vertical level). An exact match between the theoretical error and the applied error is found when each discrete value of expected SNR is investigated individually (not shown).

Figure 7 right shows the same statistics for gross errors. The notation is the same as for Gaussian er-

ror. The theoretical gross error standard deviation is equal to Rg $1/[2 \ (3)^{1/2}]$. The observation count also allows a check on the correct application of the gross error probability, since adding up the counts for both errors does match the nature run data count for each pressure level. Note the very high gross error rate for DWL in the stratosphere and upper troposphere due to the lack of backscattering particles (Fig. 7d) and the decreasing measurement error toward the lower troposphere (Fig. 7c). This behavior of the error characteristics is very realistic. The accuracy of the DWL is about 2.5 times smaller than for TEMP. However, we should note here that this number depends rather critically on the aerosol loading, which in turn is not very well known. The assumptions on the vertical aerosol distribution in this study are conservative (Stoffelen et al. 1994). Whether or not the availability of many measurements providing

TABLE 2a. Standard deviation σ of error for simulated surface observations, TEMP, PILOT, SATOB, and AIREP.

σ	pp (hPa)	u, v (m s⁻¹)	T (K)	rH (%)	Level
SYNOP	1.0	2.0	2.5	13	Surface
SHIP	1.0	2.0	2.0	13	Surface
AIREP		4.0	2.5		1000 hPa
AIREP		4.0	1.9		850 hPa
AIREP		4.0	1.7		All other levels
SATOB		1.8			1000 hPa
SATOB		1.8			850 hPa
SATOB		1.9			700 hPa
SATOB		2.1			500 hPa
SATOB		4.0			400 hPa
SATOB		4.6			300 hPa
SATOB		5.0			Above
DRIBU	1.3	2.2	2.0	13	Surface
TEMP		2.0	1.2	13	1000 hPa
TEMP		1.8	0.8	14	850 hPa

σ	pp (hPa)	u, v (m s⁻¹)	T (K)	rH (%)	Level
TEMP		1.6	0.8	16	700 hPa
TEMP		2.1	0.8	18	500 hPa
TEMP		2.6	0.8	20	400 hPa
TEMP		3.0	0.9	22	300 hPa
TEMP		3.2	1.2	24	250 hPa
TEMP		2.7	1.2	26	200 hPa
TEMP		2.4	1.0	28	150 hPa
TEMP		2.1	0.8	30	100 hPa
TEMP			0.8		70 hPa
TEMP			0.9		50 hPa
TEMP			1.0		30 hPa
TEMP			1.5		10 hPa
PILOT	same	as	for	TEMP	
PAOB	2.0				Surface

global coverage and high vertical resolution but with low accuracy is of less benefit for data assimilation than fewer accurate observations can only be investigated with extended OSSEs. This task is left to other data assimilation system developers for reasons mentioned earlier. However, a quick test assimilation of DWL data has proven beneficial over data-void areas (Haution 1995). Figure 8 shows rms statistics in σ_0 space plotted over wind speed intervals of 1 m s⁻¹. Note that the curve depicts nicely the applied error function with the minimum at $\sigma_0 = 0.05$ and 16 m s⁻¹.

e. Final database

Time series of single observation stations reveal the behavior of the nature run and the simulated value. Figure 9 shows surface pressure in Sule Skerry at 59.05°N and 4.24°W. The smooth line depicts the nature run, and the noisy line depicts the simulated surface pressure. Long spikes indicate the application of gross errors. The simulated ASCAT data were processed by the ERS1 retrieval module developed at ECMWF to

compare the retrieved with the input wind vectors and calculate the error characteristics, which validates the assumptions made on the ASCAT σ_0 error. This check was run on every day of ASCAT data to recognize any changes in the overall statistics over the period of the production.

An optimal interpolation assimilation experiment [see Lorenc (1981) and ECMWF (1992) for a description of the ECMWF data assimilation system] was run close to the starting date of the OSSE database production where the simulated observations from the database and the true state of the atmosphere are still very similar. Observation types SYNOP, AIREP, SATOB, DRIBU, TEMP, PILOT, and TOVS were extracted from the database. Two assimilations have been run at T106 resolution (note that the equivalent horizontal resolution is 125 km and different from the nature run) from a 6-h time window around 0600 UTC 5 February 1993. One experiment assimilated 6 h of original data from the GTS describing the true state of the atmosphere, and the other assimilated 6 h of simulated data from the OSSE database. The results of the assimilations are to a very high degree similar as can be seen from comparison of Fig. 10a with Fig. 10b. These validation tasks increase our confidence that the production of the database has been conducted successfully.

4. Contents of the OSSE database

The OSSE database contains simulated observations and nature run values, the latter represent the true observations before the error generation, over a 30-day period at each observation point. Tables 1a and 1b describe the parameters that are available.

TOVS brightness temperatures are provided for the High-Resolution Infrared Radiation Sounder (HIRS) channels 1–8 and 10–19, the stratospheric sounding unit (SSU) channels 1–3, and the microwave sounding unit (MSU) channels 1–4. ATOVS brightness temperatures are provided for the same HIRS channels 1–8 and 10–19 and for the Advanced Microwave Sounding Unit (AMSU)-A channels 1–15 and AMSU-B channels 16–20. These channels are part of

Table 2b. Gross error probability Pg for simulated surface observations, TEMP, PILOT, SATOB, and AIREP. Gross errors are only applied to 10% of all TEMP and PILOT reports. Ninety percent of all TEMP and PILOT reports do not contain any gross errors.

Pg	pp	u, v	T	rH	Level
SYNOP	0.015	0.020	0.020	0.040	Surface
SHIP	0.060	0.060	0.070	0.050	Surface
AIREP		0.030	0.030	·	All levels
SATOB		0.030			All levels
DRIBU	0.030	0.040	0.070	0.050	Surface
TEMP		0.150	0.150	0.500	All levels
PILOT		0.150	0.150	0.500	All levels
PAOB	0.004				Surface

the ATOVS design to allow better retrievals of atmospheric temperature and water vapor profiles from the surface up to 2 hPa. The 31 TEMP and PILOT pressure levels 10, 30, 50, 70, 90, 100, 125, 150, 175, 200, 250, 300, 335, 370, 400, 450, 500, 540, 580, 620, 660, 700, 730, 770, 810, 850, 900, 935, 970, 1000, and 1010 hPa are the standard pressure levels and additional levels, which are close to the forecast model's vertical nodes. Each of these simulated observations has an appendix that describes the history of the simulated datum. It contains the simulated true data and the applied error characteristics as described in section 2e. The five DWL instrument scenarios are

1) an 800-km orbit, with 10-J transmitted energy, a 0.95-Hz sampling frequency, and a vertical resolution of about 950 m;
2) an 800-km orbit, with 5-J transmitted energy, a 0.95-Hz sampling frequency, and a vertical resolution of about 950 m;
3) a 525-km orbit, with 10-J transmitted energy, a 0.95-Hz sampling frequency, and a vertical resolution of about 950 m; ·

4) a 525-km orbit, with 5-J transmitted energy, a 0.95-Hz sampling frequency, and a vertical resolution of about 950 m; and
5) an 800-km orbit, with 10-J transmitted energy, a 9.5 Hz sampling frequency, and a vertical resolution of about 270 m for a period of 1 day in the North Atlantic.

5. Conclusions

For the first time, a database with spatial coverage comparable to that received every day from the global observation network available via GTS and with known error characteristics derived from many years experience of analyzing these data (Lorenc et al. 1991; Gandin et al. 1993; Eyre 1990; Eyre et al. 1993; J. Eyre 1994, personal communication; Ingleby and Parrett 1994; R. Graham 1994, personal communication; ECMWF 1992) is available for OSSEs. The most important features of the database are

- multitude of conventional and satellite observational systems with realistic temporal and spatial distribution;

Table 2c. Gross error range Rg for simulated surface observations, TEMP, PILOT, SATOB, and AIREP.

Rg	pp	u, v	T	rH	Level
SYNOP	31.25	23.26	33.3	100.0	Surface
SHIP	31.25	23.26	33.3	100.0	Surface
AIREP		70.71	33.3		All levels
SATOB		25.82			All levels
DRIBU	23.26	0.04	100.0	100.0	Surface
TEMP		70.71	100.0	100.0	All levels
PILOT		70.71	100.0	100.0	All levels
PAOB	23.26				Surface

TABLE 3a. Standard deviation σ of error, gross error probability Pg, and gross error range Rg for simulated TOVS brightness temperatures.

σ			Pg			Rg	Channel
Clear sky	Partly cloudy sky	Cloudy sky	Clear sky	Partly cloudy sky	Cloudy sky		
0.70	0.70	0.70	0.010	0.010	0.010	20	HIRS-1
0.35	0.35	0.35	0.005	0.005	0.005	20	HIRS-2
0.30	0.30	0.30	0.005	0.005	0.005	20	HIRS-3
0.20	0.20		0.001	0.002		20	HIRS-4
0.30	0.30		0.002	0.004		20	HIRS-5
0.40	0.50		0.005	0.010		20	HIRS-6
0.60	0.90		0.010	0.020		20	HIRS-7
1.10	1.90		0.010	0.020		20	HIRS-8
0.80	1.30		0.010	0.020		20	HIRS-10
1.10	1.10		0.002	0.004		20	HIRS-11
1.50	1.50		0.005	0.010		20	HIRS-12
0.50	0.90		0.010	0.002		20	HIRS-13
0.35	0.60		0.005	0.010		20	HIRS-14
0.30	0.40		0.001	0.002		20	HIRS-15
0.35	0.35		0.001	0.002		20	HIRS-16
1.00	2.00		0.010	0.002		20	HIRS-17
1.00	2.00		0.010	0.020		20	HIRS-18
1.00	2.00		0.010	0.020		20	HIRS-19
1.00	1.00	1.00	0.010	0.010	0.010	20	MSU-1
0.30	0.30	0.30	0.005	0.005	0.005	20	MSU-2
0.22	0.22	0.22	0.001	0.001	0.001	20	MSU-3
0.25	0.25	0.25	0.005	0.005	0.005	20	MSU-4
0.40	0.40	0.40	0.010	0.010	0.010	20	SSU-1
1.00	1.00	1.00	0.015	0.015	0.015	20	SSU-2
1.80	1.80	1.80	0.020	0.020	0.020	20	SSU-3

TABLE 3b. Standard deviation σ of error, gross error probability Pg, and gross error range Rg for simulated ATOVS. The simulated HIRS channels on ATOVS are the same as for TOVS.

σ	Pg	Rg	Channel
1.15	0.005	20	AMSU-1
1.15	0.005	20	AMSU-2
1.15	0.005	20	AMSU-3
0.59	0.002	20	AMSU-4
0.33	0.001	20	AMSU-5
0.29	0.001	20	AMSU-6
0.27	0.001	20	AMSU-7
0.28	0.001	20	AMSU-8
0.28	0.001	20	AMSU-9
0.32	0.001	20	AMSU-10
0.32	0.002	20	AMSU-11
0.40	0.003	20	AMSU-12
0.50	0.004	20	AMSU-13
0.76	0.005	20	AMSU-14
1.14	0.005	20	AMSU-15
1.14	0.005	20	AMSU-16
1.14	0.005	20	AMSU-17
1.52	0.002	20	AMSU-18
1.05	0.002	20	AMSU-19
0.75	0.002	20	AMSU-20

FIG. 5. Standard deviation and mean of the nature run DWL wind components (m s^{-1}), relative humidity (%), cloud cover (%), and cloud liquid water (g kg^{-1}) averaged over the Northern Hemisphere north of 20°N. The left-hand axis shows vertical level in hPa, and the right axis shows the number of measurements per level. The figure refers to all data in a period of 1 day in a DWL scenario with 800-km orbit. Solid lines are standard deviations; the dotted lines are means.

- generation of a high-resolution DWL scenario with 800-km orbit and 10-J laser; and
- generation of future satellite systems ATOVS and ASCAT.

- realistic error characteristics through the inclusion of the concept of gross errors;
- simulation of a temporally consistent true atmospheric state for all meteorological parameters, including three-dimensional clouds on all model gridpoints over a period of 30 days;
- detailed simulation of four different DWL scenarios with varying orbit height and laser power;

Any data assimilation center may conduct impact assessments of DWL, ASCAT, ATOVS, or any other combination of the simulated observing systems using the database. The provision of the simulated observation together with the true observation, the standard deviation, the probability of gross error, and the gross error range can help developers of assimilation- or data handling systems to check their schemes for consistency over periods of up to 30 days.
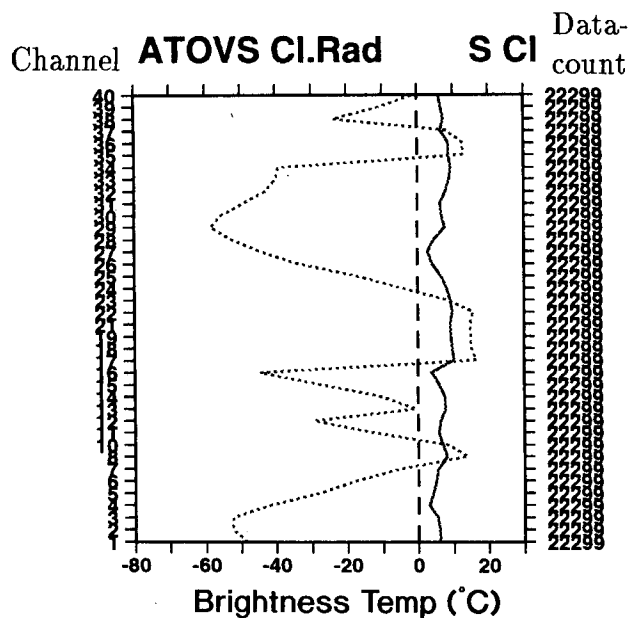
FIG. 6. Simulated clear column brightness temperatures by channel number for the ATOVS instrument mounted on the NOAA polar orbiter over sea, as they are simulated from the radiative transfer model within the IFS (Eyre 1991). The left-hand axis shows satellite channel number, and the right axis shows the number of measurements per channel. The figures refer to all data in a period of 1 day, averaged over the Northern Hemisphere north of 20°N. Solid lines are standard deviations; the dotted lines are means.
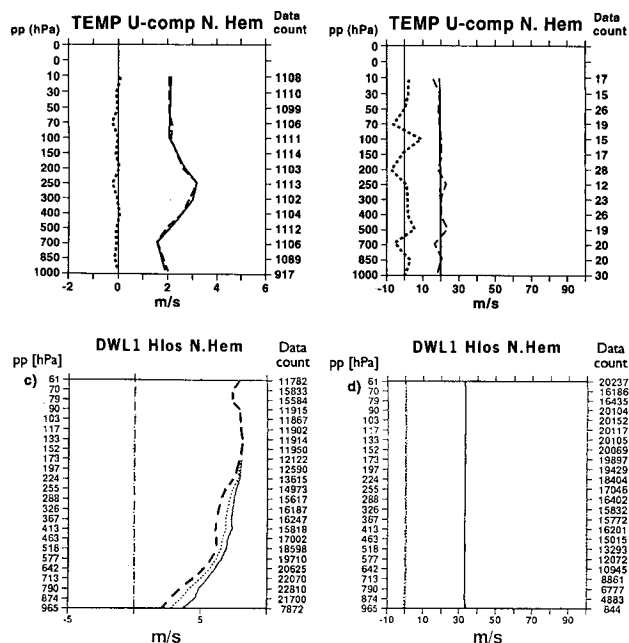

FIG. 7. Rms statistics of Gaussian and gross error of the simulated TEMPs U-wind component (a) and (b); and DWL HLOS (c) and (d). As long as the theoretical error standard deviation applied in the error model is constant with the atmospheric state, it is identical with the rms of the applied error. The mean of the applied error approaches zero when a higher number of samples is taken. The left-hand axis shows vertical level in hPa, and the right axis the number of measurements per level. The figures refer to all data in a period of 1 day, averaged over the Northern Hemisphere north of 20°N. Thin solid line: theoretical σ as used in the error model; thick-dashed line: rms of the applied error; dotted line: mean of the theoretical error; and dot–dashed line: mean of the applied error.

The well-documented error statistics and the provision of the true nature run values may aid in finding the systematic errors introduced by the data handling routines rather than the data itself.

The only data type not simulated to date is the high spectral resolution radiances from advanced infrared sounders such as the Atmospheric Infrared Sounder and the Infrared Atmospheric Sounding Interferometer. It is planned to add the latter to the database in the near future.

World Wide Web address of the ECMWF:
 http://www.ecmwf.int/
 library: http://www.ecmwf.int/library/home.html
 miscellaneous publications: http://www.ecmwf.int/library/misc.html

Appendix: The OSSE database files

The whole of the OSSE database resides at ECMWF. The file-naming convention is bufr-$TYPE$DATE, where $TYPE can be synop, tovs, atovs, dwl1, dwl2, and scatt (see Table A1). $DATE may be any date in the form YYMMDD from 930205 to 930306. Individual files can be requested for a particular date or the whole period.

The total volume of this database is approximately 2.9 GB, split across 241 files. Each BuFR report contains the simulated data, the nature run values, the standard deviation of error σ, the probability of a gross error Pg, the gross error range Rg, and a gross

TABLE A1. ECMWF OSSE database file-naming convention and sizes. Any date between 930205 and 930306 may be used for *.

File	Contents	Size (MB)
bufrsynop*	SYNOP, AIREP, SATOB, DRIBU, TEMP, PILOT, and PAOB	3
bufrtovs*	HIRS 1-8 and 10-19, SSU 1-3 and MSU 1-4	7
bufratovs*	HIRS 1-8 and 10-19, AMSU-A 1-15 and AMSU-B 1-5	9
bufrscatt*	ASCAT double swath σ_0	5
buf1dwl1*	High-orbit, high-energy DWL	18
buf1dwl2*	Low-orbit, high-energy DWL	18
buf2dwl1*	High-orbit, low-energy DWL	18
buf2dwl2*	Low-orbit, low-energy DWL	18
buf3dwl1930206	High-resolution, high-orbit DWL	8

error tag. BuFR decoding routines are needed and these can be provided.

The nature run model fields are stored in the ECMWF Meteorological Archive and Retrieval System (MARS) as one 30-day forecast, starting 93020500, including all model fields every 6 h over the whole period. A description of the nature run model fields stored in MARS is given in Stoffelen et al. (1994). Access to the database can be granted by ESA. Applications should be sent to

ESA/ESTEC
Director of the Earth Science Division
Dr. C. J. Readings
Postbus 299
NL-2200 AG Noordwijk
Netherlands

When access is granted, contact R. Saunders at ECMWF (r.saunders @ecmwf.int) for further details of the data distribution procedure.
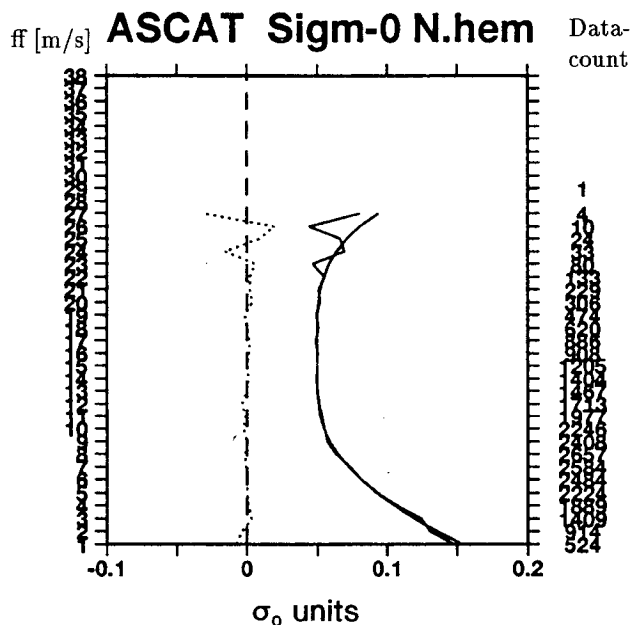
ff [m/s] **ASCAT Sigm-0 N.hem** Data-count

σ₀ units

FIG. 8. Rms statistics of the simulated ASCATs. Rms statistics in σ_0 space plotted over wind speed intervals in m s^{-1}. The curve depicts nicely the applied error function $e = \sigma_0(zkp^2 + 0.05^2)^{1/2}$, zkp = 0.000644 (ff− 16)^2 with the minimum at $\sigma_0 = 0.05$ and 16 m s^{-1}. The left-hand axis shows wind speed in m s^{-1}, and the right axis shows the number of measurements per 1 m s^{-1} interval. The figures refer to all data in a period of one day, averaged over the Northern Hemisphere north of 20°N. Solid: theoretical and simulated σ; dotted: bias of simulated data per wind speed intervals of 1 m s^{-1}.

References

Andersson, E., A. Hollingsworth, G. Kelly, P. Lönnberg, J. Pailleux, and Z. Zhang, 1991: Global observing system experiments on operational statistical retrievals of satellite sounding data. Mon. Wea. Rev., 119, 1851–1864.

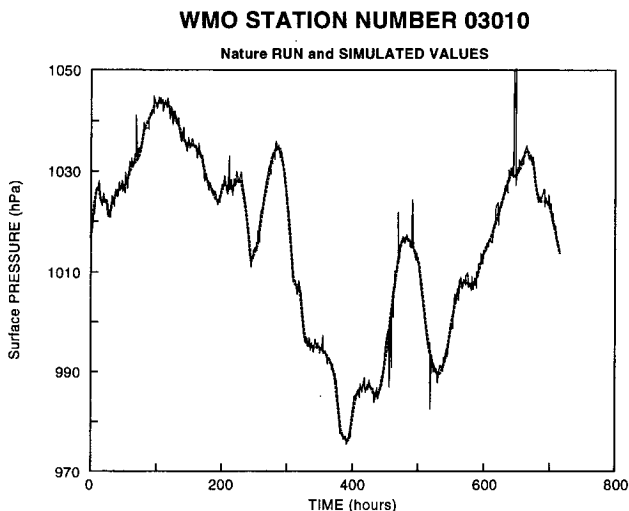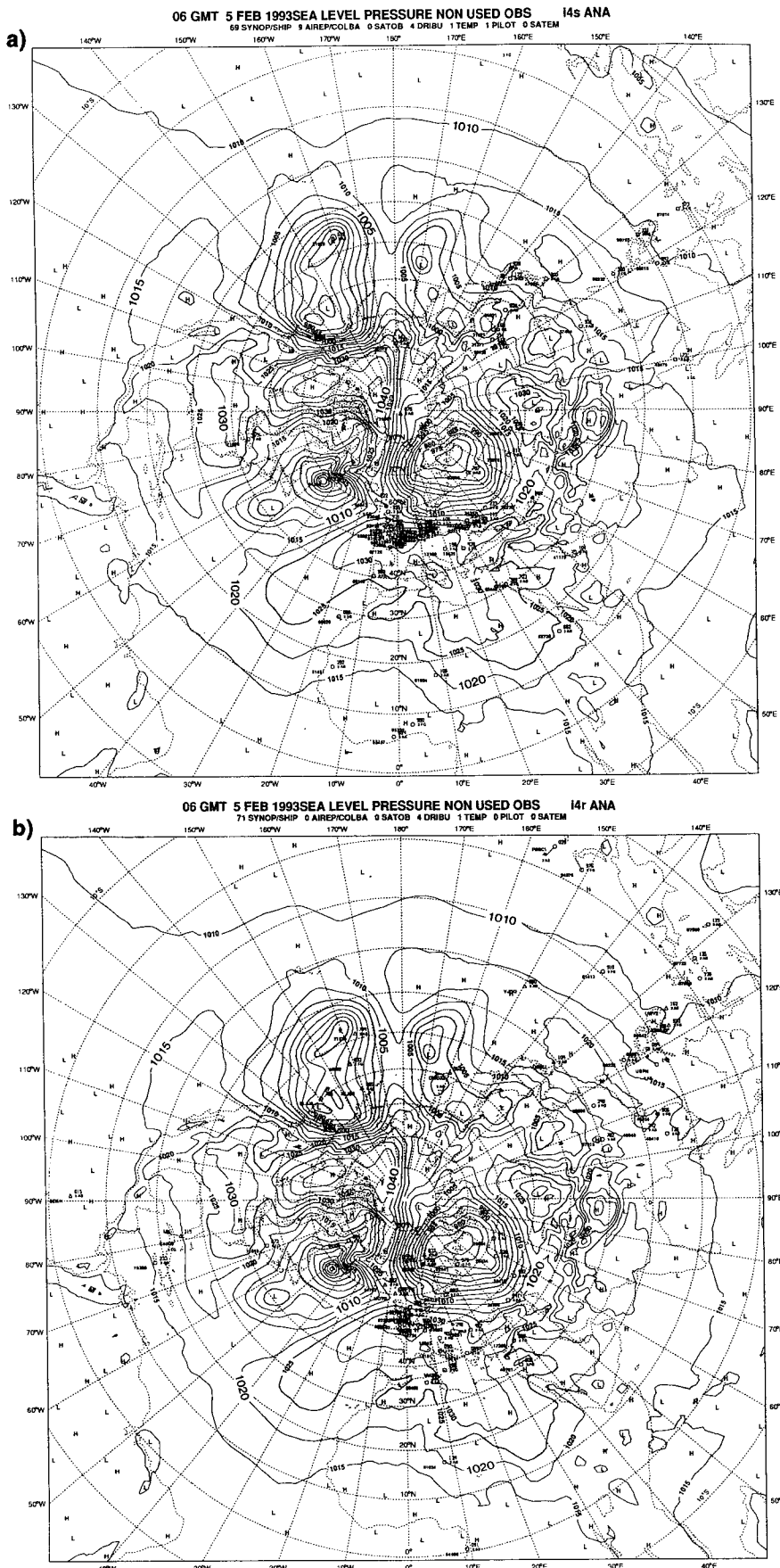**WMO STATION NUMBER 03010**

Nature RUN and SIMULATED VALUES

FIG. 9. Surface pressure in Sule Skerry at 59.05°N and 4.24°W. The smooth thick block-hashed line depicts the nature run, and the noisy line depicts the simulated surface pressure. Long spikes indicate the application of gross errors.

06 GMT 5 FEB 1993SEA LEVEL PRESSURE NON USED OBS   I4s ANA
69 SYNOP/SHIP 9 AIREP/COLBA 0 SATOB 4 DRIBU 1 TEMP 1 PILOT 0 SATEM

a)

06 GMT 5 FEB 1993SEA LEVEL PRESSURE NON USED OBS   I4r ANA
71 SYNOP/SHIP 0 AIREP/COLBA 0 SATOB 4 DRIBU 1 TEMP 0 PILOT 0 SATEM

b)

Arnold, C. P., and C. H. Dey, 1986: Observing-system simulation experiments: Past, present and future. *Bull. Amer. Meteor. Soc.*, **67**, 687–695.

Baker, W. E., and Coauthors, 1995: Lidar-measured winds from space: A key component for weather and climate prediction. *Bull. Amer. Meteor. Soc.*, **76**, 869–888.

Becker, B. D., and H. Roquet, 1995: Extension of the OSSE database to scatterometer and ATOVS data, Part II. The European Centre for Medium-Range Weather Forecasts (ECMWF) Synthetic Observation Production Suite. ESA contract rep. 10921/94/NL/CN, 59 pp. [Available from ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom.]

Betout, P., D. Burridge, and C. Werner, 1989: Döppler Lidar Working Group Rep., ESA SP-1112, 45 pp. [Available from British Lib. Doc. Supply Centre, Customer Service, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, United Kingdom.]

Carr, F., M. K. Ramamurthy, D. J. Rusk, and G. P. Lou, 1993: Observing system experiments: Relative model response to various FGGE datasets in the Tropics. *Mon. Wea. Rev.*, **121**, 3106–3122.

Courtier, P., C. Freydier, J. F. Geleyn, F. Rabier, and M. Rochas, 1991: The Arpege project at Météo-France. *ECMWF Seminar Proceedings, Numerical Methods in Atmospheric Models,* Vol. II, ECMWF 193–231.

——, P. Gauthier, F. Rabier, P. Flammant, A. Dabas, and F. Lieutaud, 1992: Study of preparation for the use of Döppler wind lidar information in meteorological assimilation systems. Final Rep. on ESA Study Contract 8850/90/HGE-I, CNRM, Météo-France, Paris, France. [Available from ESA, Earth

Fig. 10. (a) Analyzed simulated and (b) analyzed original sea level pressure of the Northern Hemisphere for a 6-h period around 0600 UTC on 5 February 1993. The results of the assimilations of true data from the GTS and simulated OSSE data are to a very high degree similar.

Science Division, Postbus 299, NL-2200 AG Noordwijk, the Netherlands.]

ECMWF, 1992: ECMWF data assimilation scientific documentation. Meteorological Bulletin M1.5/1, 2.32 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.]

Eyre, J., 1990: The information content of data from satellite systems: A simulation study. *Quart. J. Roy. Meteor. Soc.*, **116**, 401–434.

——, 1991: A fast radiative transfer model for satellite sounding systems. ECMWF Tech. Memo. 176, 28 pp. [Available from ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom.]

——, G. A. Kelly, A. P. McNally, E. Anderson, and A. Persson, 1993: Assimilation of TOVS radiance information through one-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **119**, 1427–1463.

Gandin, L. S., L. L. Morone, and W. G. Collins, 1993: Two years of comprehensive hydrostatic quality control at the National Meteorological Center. *Wea. Forecasting*, **8**, 57–72.

Haution, P., 1995: Impact potentiel des mesures de vent par lidar Doppler sur la prévision numérique en météorologie. Thesis, Centre de Géostatistique, Ecole de Mines de Paris. [Available from Centre de Géostatistique, Ecole de Mines de Paris, 35 Rue Saint-Honore, 77305 Fontaineblue, Paris, France.]

Hollingsworth, A., and P. Lönnberg, 1989: The verification of objective analyses: Diagnostic of analysis system performance. *Meteor. Atmos. Phys.*, **40**, 3–27.

Hortal, M., and A. J. Simmons, 1991: Use of reduced Gaussian grids in spectral models. *Mon. Wea. Rev.*, **119**, 1057–1074.

Ingleby, N. B., and C. A. Parrett, 1994: Quality control of atmospheric data. U.K. Meteorological Office, Unified Model Documentation Paper 32. [Available from NWP Division, U.K. Meteorological Office, London Road, Bracknell, Berkshire RG12 2SZ, United Kingdom.]

Jarraud, M., J. Goasand, and C. Deyts, 1989: Prediction of an exceptional storm over France and southern England (15–16 October 1987). *Wea. Forecasting*, **4**, 517–536.

Lorenc, A., 1981: A global three-dimensional multivariate statistical analysis scheme. *Mon. Wea. Rev.*, **109**, 701–721.

——, R. Graham, I. Dharssi, B. McPherson, N. B. Ingleby, and R. W. Lunnon, 1991: Study of preparation for use of Doppler wind lidar information in meteorological assimilation systems. Final Rep. on ESA Study Contract 8850/90/HGE-I. [Available from NWP Division, U.K. Meteorological Office, London Road, Bracknell, Berkshire RG12 2SZ, United Kingdom.]

Miller, M. J., T. N. Palmer, and R. Swinbank, 1989: Parameterization and influence of subgrid-scale orography in general circulation and numerical weather prediction models. *Meteor. Atmos. Phys.*, **40**, 84–109.

Morcrette, J.-J., 1990: Impact of changes in the radiation transfer parameterization plus cloud optical properties in the ECMWF model. *Mon. Wea. Rev.*, **118**, 847–873.

Ritchie, H., C. Temperton, A. Simmons, M. Hortal, T. Davies, D. Dent, and M. Hamrud, 1995: Implementation of the semi-Langrangian method in a high-resolution version of the ECMWF forecast model. *Mon. Wea. Rev.*, **123**, 489–514.

Rohaly, G. D., and T. N. Krishnamurti, 1993: An observing system simulation experiment for the Laser Atmospheric Wind Sounder (LAWS). *J. Appl. Meteor.*, **32**, 1453–1471.

Roquet, H., B. D. Becker, and R. Saunders, 1995: Extension of the OSSE database to scatterometer and ATOVS data, Part I. ESA Contract Rep. 10921/94/NL/CN, 46 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.]

Simmons, A. J., and R. Strüfing, 1981: An energy and angular-momentum conserving scheme, hybrid coordinates and medium-range weather prediction. ECMWF Tech. Rep. 28, 68 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.]

Stoffelen, A., and D. Anderson, 1995: The ECMWF contribution to the characterization, interpretation, calibration and validation of ERS-1 scatterometer backscatter measurements and winds, and their use in numerical weather prediction model. ESA Contract Rep. 9097/90/NL/BI, 92 pp. [Available from ESA, Earth Science Division, Postbus 299, NL-2200 AG Noordwijk, the Netherlands.]

——, B. D. Becker, J. Eyre, and H. Roquet, 1994: Theoretical studies of the impact of Doppler wind lidar data, preparation of a database. ESA Contract Rep. AO/1-2345/90/HGE-1, 100 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.]

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parametrization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800.

Viterbo, P., and A. C. M. Beljaars, 1995: An improved land surface parametrization scheme in the ECMWF model and its validation. *J. Climate*, **8**, 2716–2748.