ESA CONTRACT REPORT

European Space Agency Contract Report

Theoretical Studies of the Impact of Doppler Wind LIDAR Data.

Preparation of a Data Base

Authors: Ad Stoffelen, Bernd Becker, John Eyre, Hervé Roquet*

ESA Contract No: AO/1-2345/90/HGE-1

ESA Study Manager: E Oriol

* *Current affiliation KNMI, de Bilt, The Netherlands*

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

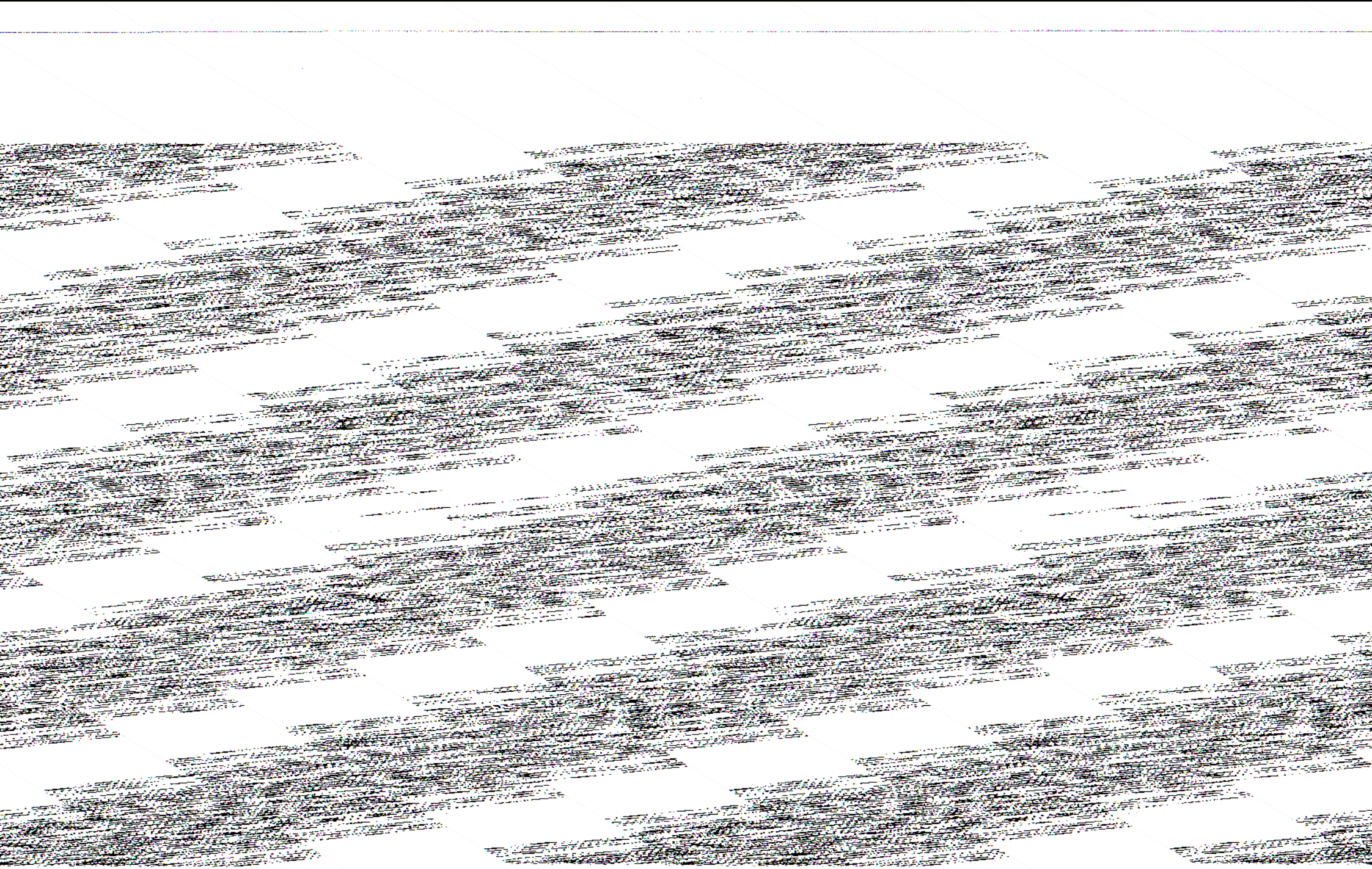
EXECUTIVE SUMMARY

For the purposes of Numerical Weather Prediction (NWP), the current network of conventional observations suffers from important weaknesses, especially over oceans. This is the reason why, in the last two decades, considerable effort has been put into complementary space-based remote sensing systems. However, a lot of problems arise when using them in NWP: some of them have a low vertical resolution and an indirect link with atmospheric state variables (e.g. TOVS data), some others are single level data (e.g. ERS-1 scatterometer data). Doppler Wind Lidar (DWL) data seem very promising because wind is the most important parameter in the Tropics at synoptic scales and everywhere at smaller horizontal scales. However, the contribution of DWL data to NWP will strongly depend on their amount, quality and distribution. In this respect, the data base produced by this study will be very useful, since it includes the simulation of 4 different DWL scenarios (2 different orbital heights and 2 different instruments) over the globe and a period of one month. A high resolution DWL data set has also been generated for one day over the Northern Atlantic (temporal sampling pulse rate of 9.5 Hz).

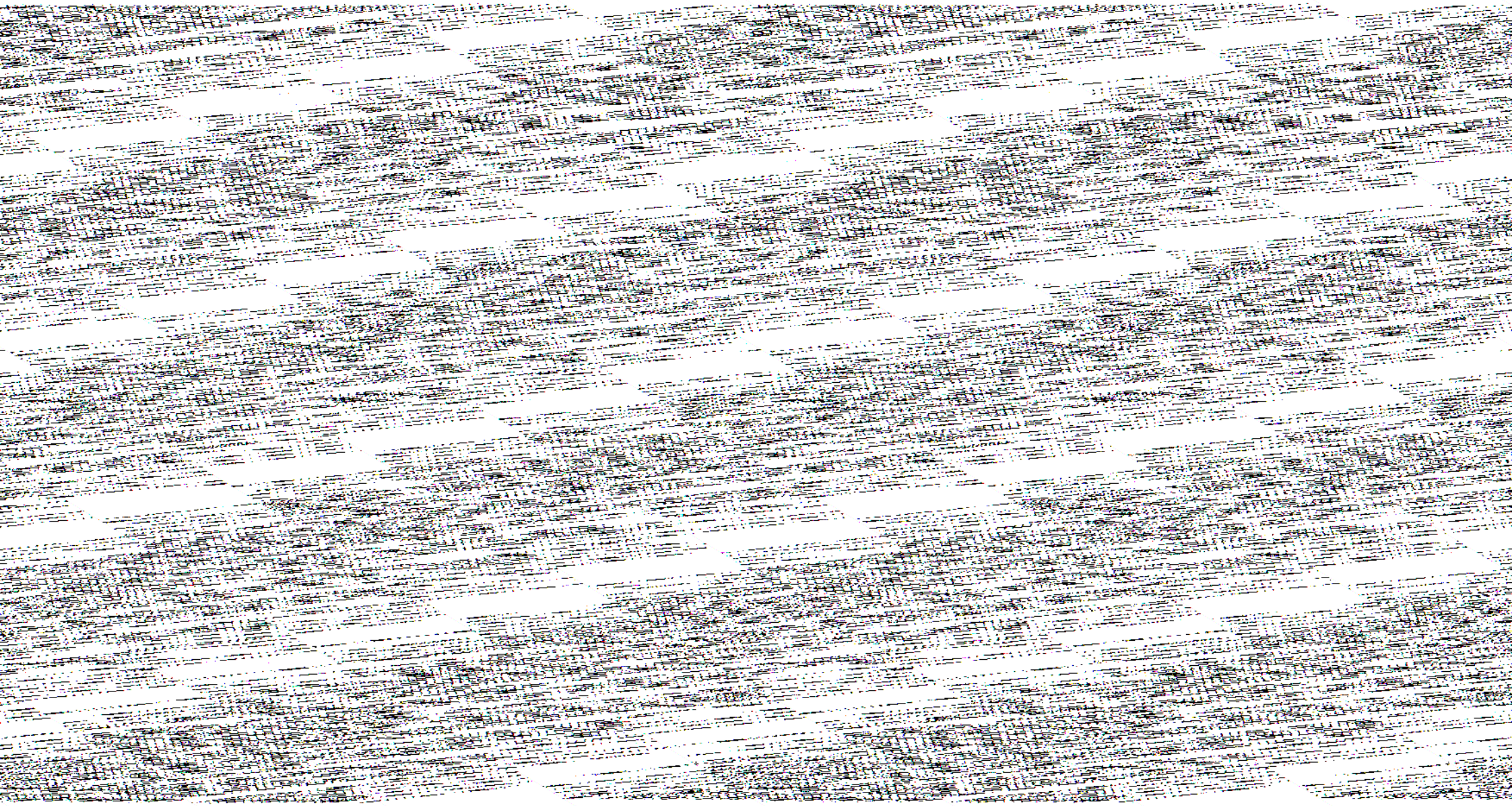
In the data base, conventional observations (land SYNOP, SHIP, BUOY, AIREP, TEMP, PILOT, PAOB and SATOB) were simulated using the true data distribution of February 1993. TEMP and PILOT soundings were generated on 31 pressure levels, to take advantage of the forecast model vertical resolution. TOVS and DWL data were generated using an orbit simulator. In the TOVS case, clear-column brightness temperatures were simulated for a two-satellite system, imitating the characteristics of the 120 km products generated and distributed operationally by NOAA/NESDIS. In the DWL case, measurements were simulated for a CO_2 laser instrument at a wavelength of λ = 9.1 μm, with a conical scan at 45° incidence angle and a 10s period. The different scenarios correspond to an instrument with a 10 J energy, which corresponds to state-of-the-art technology, and a descoped instrument with a 5J energy, for orbital heights of 525 km and 800 km. Neglecting the vertical component of the wind, we computed Horizontal Line Of Sight (HLOS) winds on equivalent pressure levels, corresponding to a vertical resolution of ~ 272 m. For the global data sets, the sampling pulse rate was set to 0.95 Hz, assuming that 10 individual measurements were made at each location, and on the vertical every 3 range gates were averaged.

The nature run, i.e. the 30-day numerical weather forecast used to generate the atmospheric fields, was performed with ECMWF's Integrated Forecast System (IFS) model, with a T213 spectral truncation and 31 levels on the vertical. Although its initial conditions were taken from the operational analysis on 5 February 1993 at 00 UT, the nature run differs substantially from reality from about day 7 onwards. The period from 5 February to 7 March 1993 was selected as a Northern Hemisphere winter case, which is not unfavourable in terms of DWL soundings in meteorologically active and interesting situations. Concerning simulated cloud characteristics, which are of great importance for DWL data, previous studies showed that the ECMWF operational cloud scheme gives a good synoptic description of the cloud amount, its main weakness being an underestimation of high level clouds water/ice content.

For the generation of the data base, IFS was run in periods of one day starting at 00 UT, and in a mode which enables the forecast fields to be compared with a set of observations distributed in space and time. These observations were stored in a format called Central Memory Arrays (CMA) format (~ 3 Gbytes per day). At each data location, IFS performed a horizontal interpolation resulting in vertical profiles of horizontal wind, geopotential, temperature and relative humidity. Then, for upper-air observations, a vertical interpolation was done to the corresponding pressure level, and an appropriate interpolation was applied for surface observations. For cloud variables, the closest model point to the observation location was selected.



CONTENTS



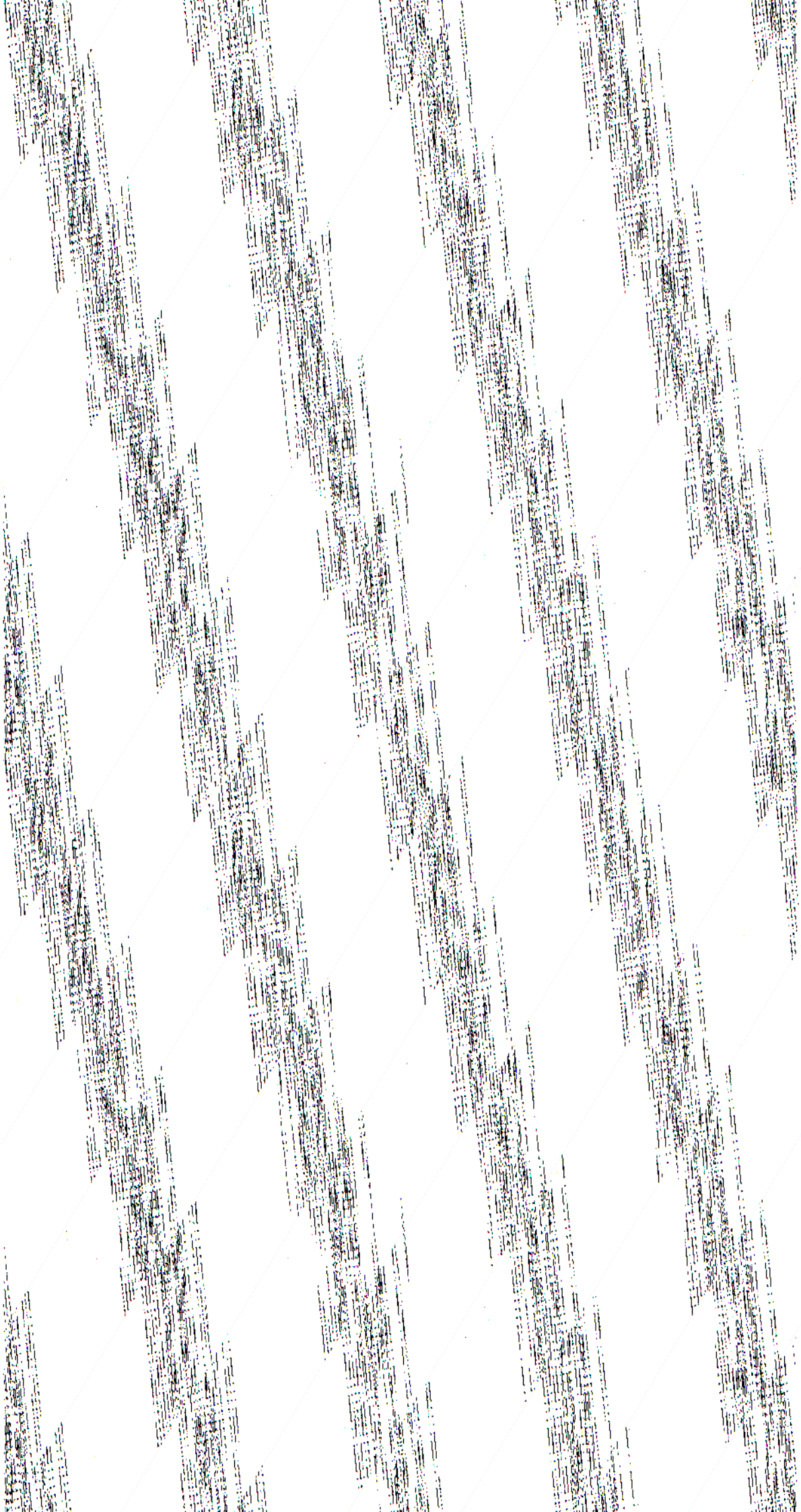
the poleward moisture transport, may therefore bias the result. A better estimate may be derived by combining these data optimally with other sources of information in a 4D NWP analysis system (*Lorenc et al*, 1991).

Because of its advanced technological requirements, a space-based system will be very costly, and therefore an assessment of its potential for NWP and climate studies is necessary (*Betout et al*, 1989). Some preliminary studies indicated that the DWL could substantially contribute to the accuracy of meteorological analysis in the Southern Hemisphere (SH) (*Courtier et al*, 1992), but this contribution will rely on the amount, quality and distribution of the LOS winds. These aspects may be investigated by realistically simulating these observations. Subsequently, Observing System Simulation Experiments (OSSEs) may further assess the potential of DWL winds for NWP, and guide the choice of orbital and instrumental parameters.

Since a threshold in signal to noise ratio (***SNR***) exists below which the signal processing collapses, it is probably better to have fewer shots with high power, compared to many shots with low power (*Courtier et al*, 1992). Further, the orbital height of the platform on which the instrument is mounted will strongly influence the number of data with sufficient ***SNR***. On the other hand, a 525 km orbital height will produce significant data gaps in the Tropics (*Courtier et al*, 1992). The DWL will not provide wind vector measurements, but LOS wind components. This is no problem for NWP data assimilation and therefore does not constrain shot management (*Lorenc et al*, 1991). In this study four different DWL orbit and instrument scenarios were simulated over a period of one month and the full globe to assess the sensitivity to a change in orbit height and emitted signal strength. The simulation was done at a temporal resolution of ~1 Hz, which results in a sampling on the Earth's surface compatible with what the ECMWF model at spectral truncation T213 can resolve. For a period of one day in the North Atlantic a high temporal resolution (~10 Hz) was used, and subgrid scale cloud effects were simulated in a statistical manner.

In preparation for OSSEs to be performed by major NWP centres, we prepared simulated data sets at ECMWF for typical distributions of all currently available GTS data and for DWL as discussed in section 2.

To perform the simulations we used the Integrated Forecasting System (IFS) as operational since 2 March 1994 at ECMWF. In this system a mode exists whereby a forecast can be run and compared simultaneously to a set of observations. The meteorological conditions at the starting date and time



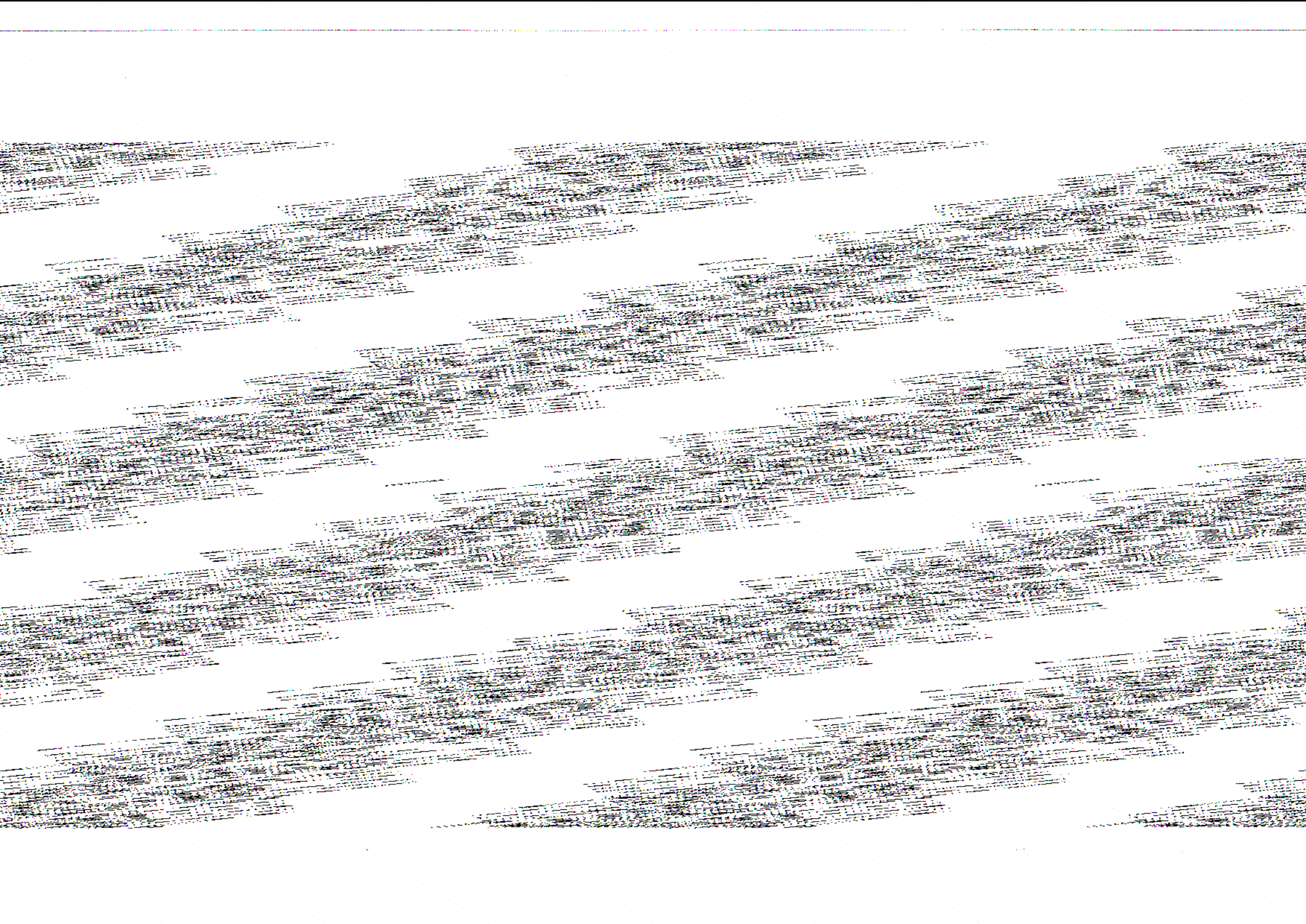
2. CONTENTS OF THE DATA BASE

As a reference for the data distribution in time and space, we took the observational reports for the period of the "nature" run as received by GTS and presented at the time to the ECMWF operational analysis. The data distribution for the satellite-based vertical sounders was generated with an orbit simulator. The reports are categorised as:

land SYNOP	surface reports over land of pressure, wind, temperature and humidity,
SHIP	surface reports over sea of pressure, wind, temperature and humidity,
BUOY	drifting and moored buoy reports of pressure, wind, temperature and humidity,
AIREP	single level upper air reports from aircraft or balloon of wind and temperature,
TEMP	vertical sounding reports of height, wind, temperature and humidity,
PILOT	vertical sounding reports with only wind,
PAOB	bogus surface pressure observations derived from imagery and ancillary information,
SATOB	cloud-tracked winds,
TOVS	NOAA satellite-based vertical soundings of radiances, and
DWL	satellite-based vertical soundings of LOS wind.

Not all stations report all variables, and transmission losses may also reduce the data volume. We took this into account when we simulated the reports. For the conventional data we ignored the station blacklists used in daily ECMWF operations, resulting in a slightly optimistic conventional data coverage. For TEMPs and PILOTs we set the number of vertical significant levels to 31 to be compatible with the vertical resolution of the forecast model. All the simulated observed values are at the correct observation time to within 30 minutes, except for the cloud variables which, due to the frequency of calling the cloud scheme, are lagged by an amount varying between 30 minutes and 3½ hours. The synoptic distribution of cloud is generally fairly constant over a period of 3 hours, except perhaps in the tropics over land during day-time.

In this section we describe the characteristics of these data types in more detail. However, the error characteristics as used in the simulation are presented in section 4.3. For convenience we discuss PAOB and SATOB observations under conventional data.



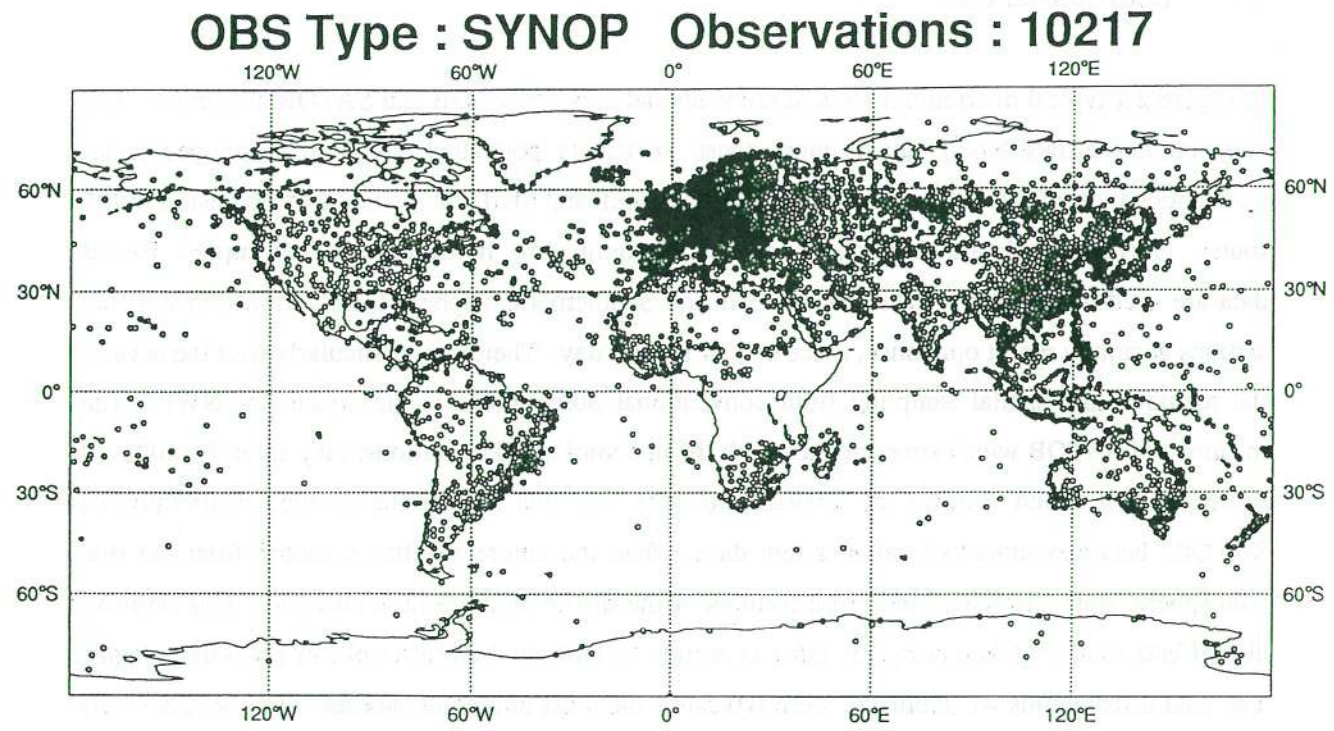
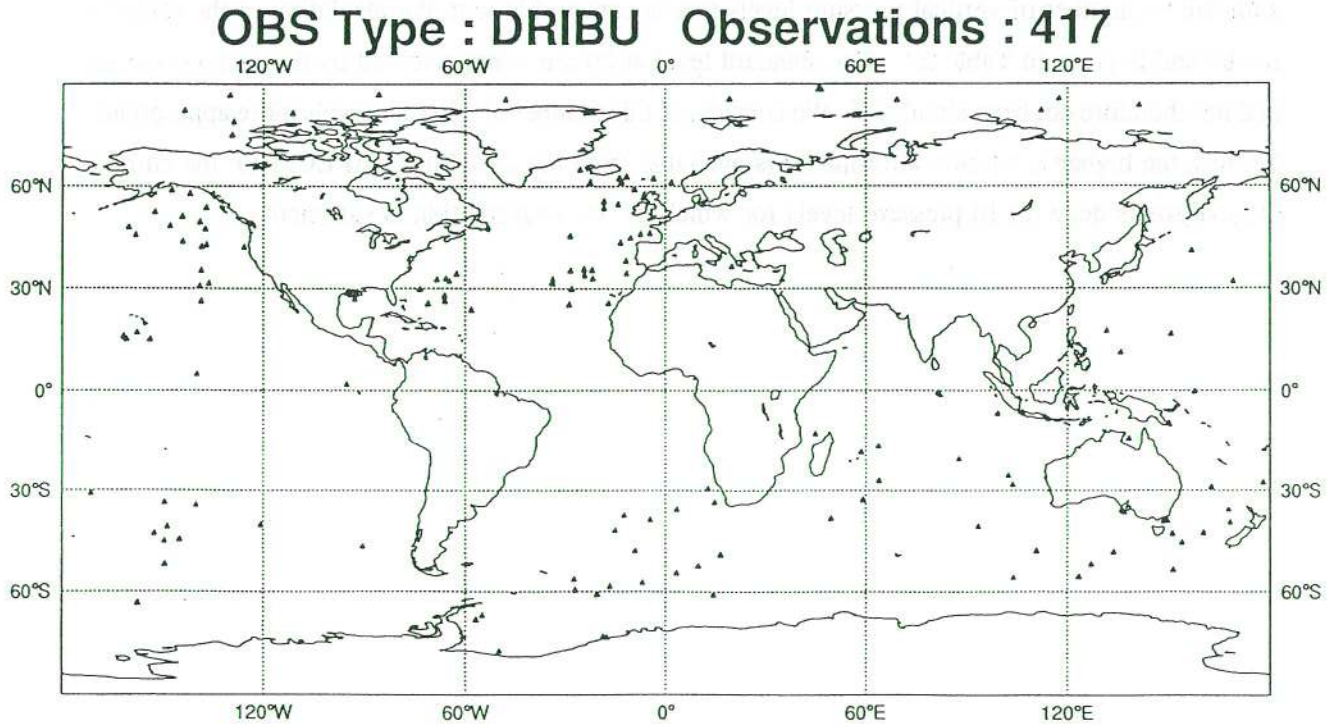
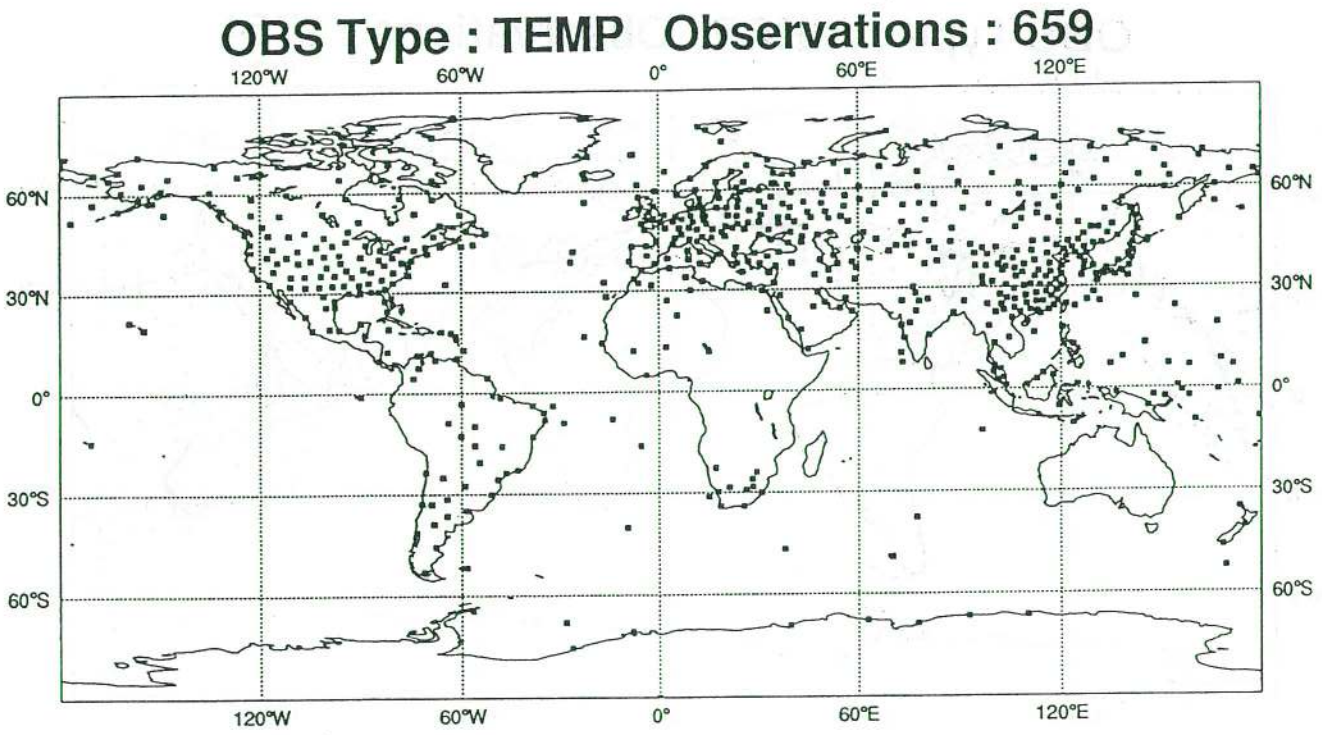


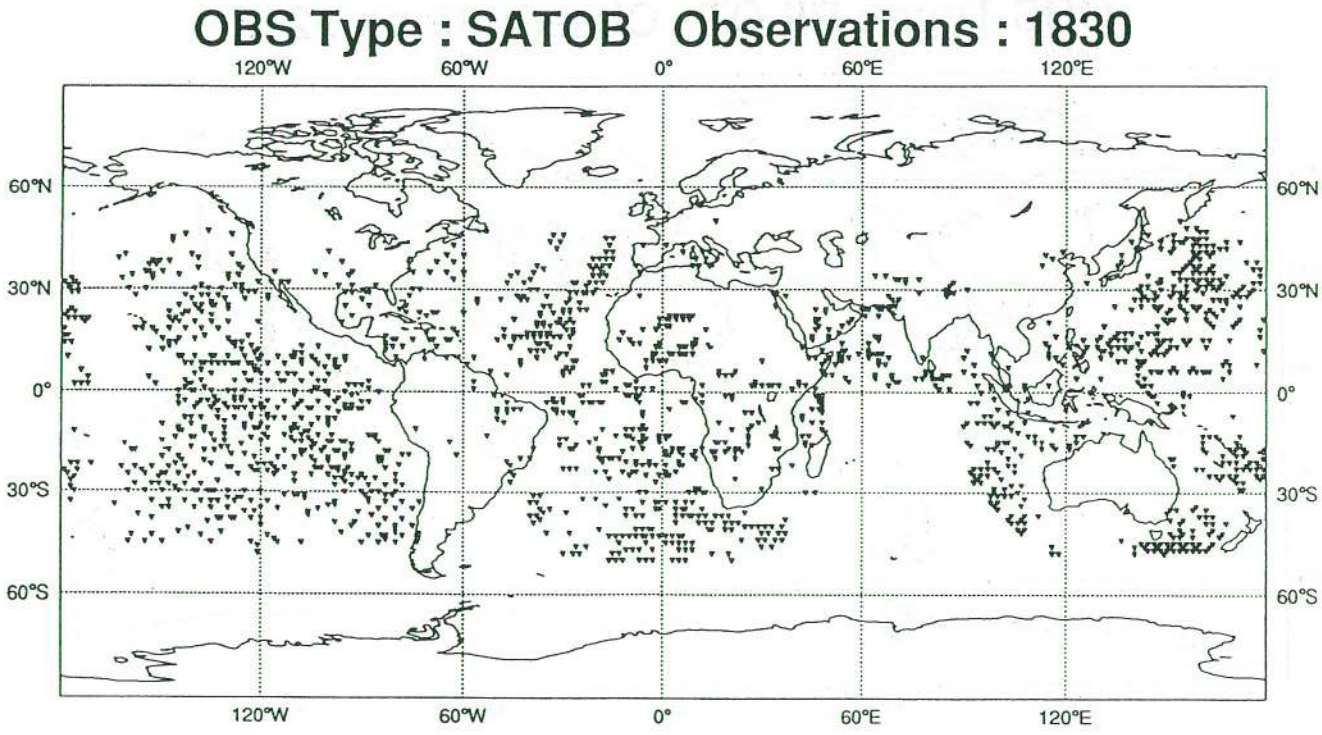
Fig 2.1:(a) Acquired distribution of observations from 09:00 UT until 15:00 UT 5/2/93 for land SYNOP and SHIP. A similar distribution is acquired at 00 UT, 06 UT and 18 UT.



(b) As (a), but for BUOY reports.



(e) As (a), but for TEMP reports.



(f) As (a), but for SATOB reports.

Table 2.1: Pressure levels for simulated TEMPs and PILOTs (column 1), reference pressure levels for the 31 level forecast model (column 2), and reference height levels for the ECMWF model. The **marked** levels are the standard levels used in the ECMWF OI analysis.

#	P (hPa) SONDE	P (hPa) L31	Z (km) L31
1	**10**	10	31.0
2	**30**	30	24.5
3	**50**	50	20.5
4	**70**	70	18.2
5	90	90	17.0
6	**100**	111	15.5
7	125	132	14.2
8	**150**	156	13.2
9	175	181	12.5
10	**200**	209	11.5
11	**250**	238	10.5
12	**300**	270	9.7
13	335	304	9.0
14	370	340	8.2
15	**400**	378	7.5
16	450	417	6.7
17	**500**	458	6.0
18	540	500	5.2
19	580	543	4.7
20	620	588	4.2
21	660	633	3.7
22	**700**	680	3.2
23	730	726	2.7
24	770	772	2.2
25	810	818	1.7
26	**850**	861	1.3
27	900	903	1.0
28	935	940	0.7
29	970	972	0.4
30	**1000**	996	0.15
31	1010	1009	0.02

2.2 TOVS

Figure 2.2 shows a distribution plot of TOVS profiles from the two NOAA satellites for a 6 hour period. Here temporal and spatial coverage is much more favourable than for conventional observation types, but the difficulty lies in their interpretation for NWP model use. In particular, the radiances measured by TOVS represent information from a thick vertical slice through the atmosphere; they yield about 7 pieces of independent information on the temperature between the surface and 30 hPa.

TOVS clear-column brightness temperatures have been simulated for a two-satellite system similar to the present operational configuration of NOAA satellites. The density and coverage has been simulated to imitate the characteristics of the so-called "120 km BUFR TOVS" data set, generated and distributed operationally by NOAA/NESDIS. In reality, clear-column brightness temperatures are derived by NESDIS through one of three "cloud-clearing" routes and, depending on the route, they are labelled as "clear", "partly cloudy" or "cloudy". The probability that the processing follows a given route is strongly dependent on the cloud conditions in the field of view of the HIRS instrument. We have imitated this process using a decision algorithm which takes account of the total cloud cover of the nature run at each TOVS location. This then determines the channels for which clear-column brightness temperatures are simulated and the error statistics assigned to them.

2.3 DWL

A Doppler wind lidar measurement relies on the Doppler frequency shift. When an optical wave is backscattered by moving aerosol particles, an observer will note a change in frequency of the optical wave. When we assume that the aerosol particles are moving with the wind speed, the Doppler shift in frequency Δf can be written as:

$$\Delta f = - 2 V_r / \lambda$$

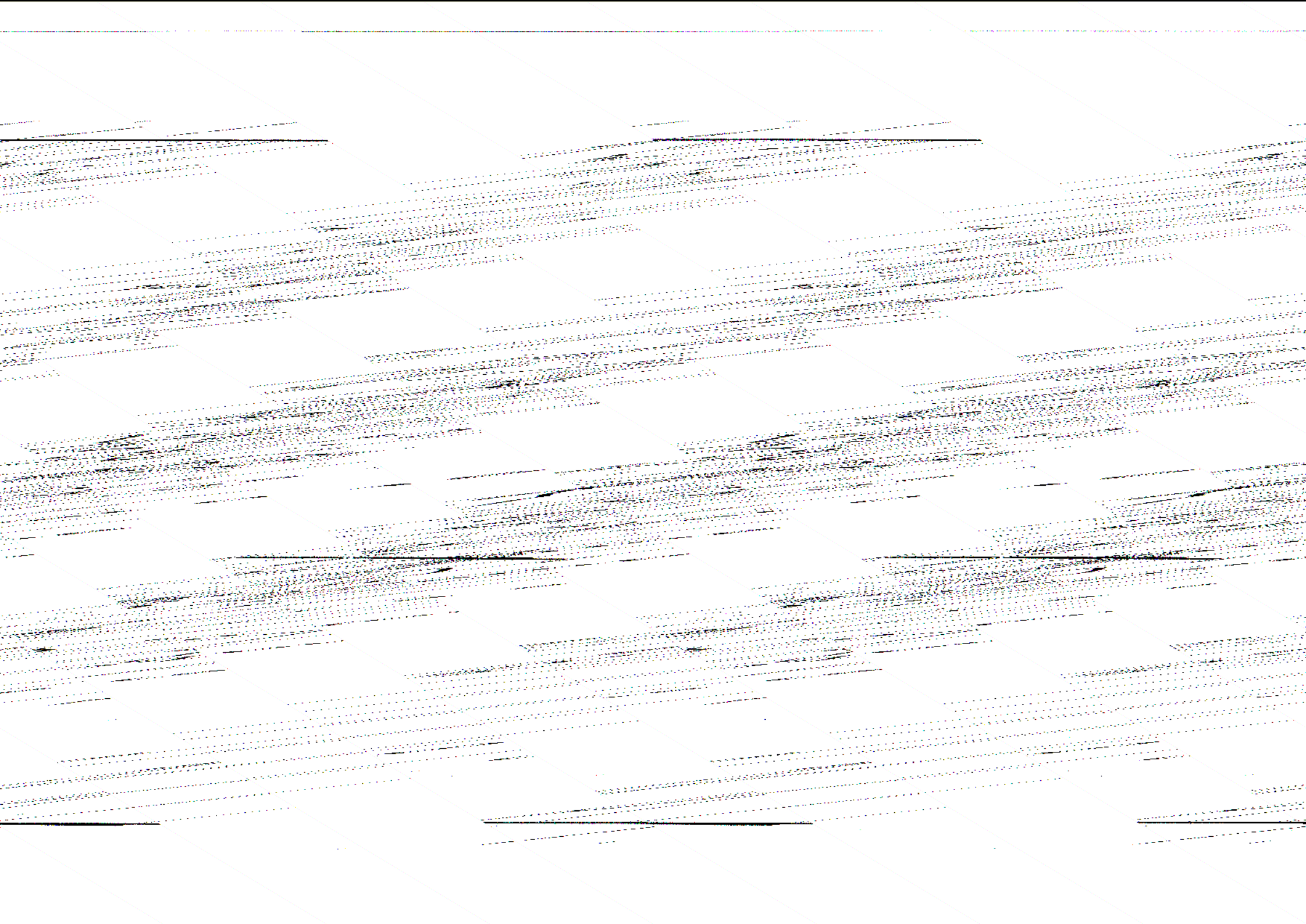
where V_r is the LOS wind speed. In this report we assumed a CO_2 laser technology instrument with a wavelength of $\lambda = 9.1$ μm. We consider this technology the most mature, and therefore most likely to be available for spaceborne application in the not too distant future. We followed the LMD laser instrument model to describe the detection of the Doppler shift (see section 4.3.3). The ***SNR*** of the received signal will depend on atmospheric backscatter properties and optical transmission in clear air and in cloud. Furthermore, the signal has a random nature due to the speckle effect, which implies that the error characteristics will be poorly described by a general error distribution, which would be desirable for data assimilation purposes. This makes the DWL simulated winds different from the other simulated observations, as would be the case in reality.

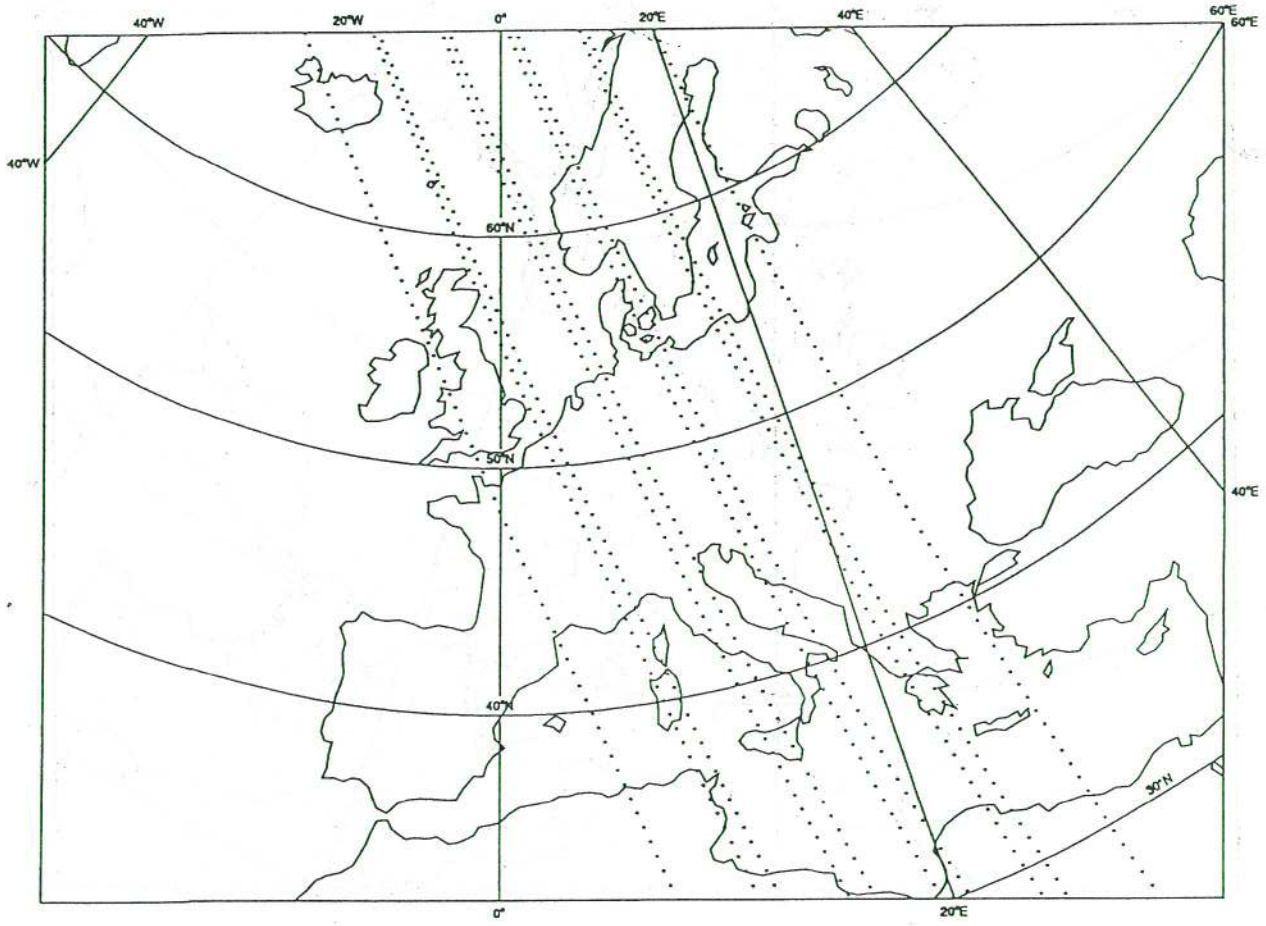
We assumed an instrument performing a conical scanning with a scan period of 10 s, a scan angle of 45° and a cosine distributed pulse rate with average frequency of f = 9.5 Hz. In Figure 2.3a we show the spatial distribution of DWL LOS winds at the Earth's surface for an orbit of 800 km. It can be noted that the sampling distance within the DWL swath is very high, i.e. on average 32 km. This scale cannot be resolved by the ECMWF model. The horizontal sampling in the

simulation was done at a temporal sampling pulse rate of f_s = 0.95 Hz which results in an average sampling distance of 100 km on the Earth's surface, which is compatible with what the ECMWF model at spectral truncation T213 can resolve (see Figure 2.3b). For an orbit of 525 km, the average sampling distance would be about 80 km. In the LOS wind simulation we subsequently assume that 10 individual measurements are made at each location (see section 4.3.3). In Figure 2.3c we show the data distribution for f = 1 Hz containing roughly the same number of data points, but with a much more unfavourable distribution than for f = 0.95 Hz. For this reason the latter frequency was used.

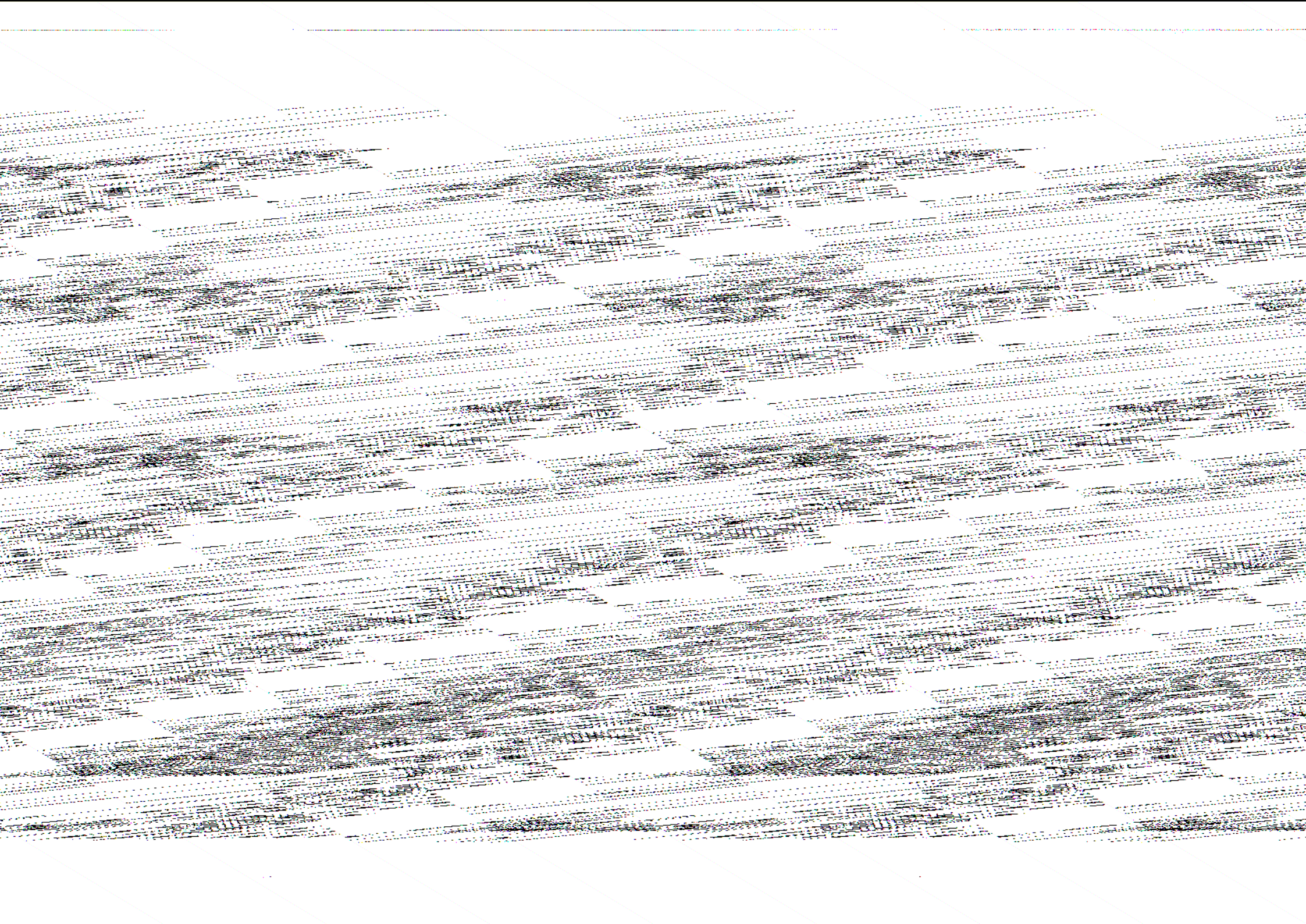
For a period of one day in the North Atlantic the high temporal resolution of 9.5 Hz was used. Backscattering is a random process (due to the speckle effect) and the extraction of reliable information from such measurements is rather complicated. With the high resolution simulations this process of information extraction can be studied.

Since a threshold in ***SNR*** exists below which the signal processing collapses, it is probably better to have fewer shots with high power, rather than many shots with low power (*Courtier et al*, 1992). Furthermore, the orbital height of the platform on which the instrument is mounted will strongly influence the number of data with sufficient ***SNR***. On the other hand, a 525 km orbital height will produce significant data gaps in the Tropics (*Courtier et al*, 1992). The DWL will not provide wind vector measurements, but LOS wind components. This is no problem for NWP data assimilation, and therefore does not constrain shot management (*Lorenc et al*, 1991). In this study four different DWL orbit and instrument scenarios were simulated over a period of one month and the full globe to assess the sensitivity to a change in orbit height and emitted signal strength. The parameter values used are given in Table 2.2.



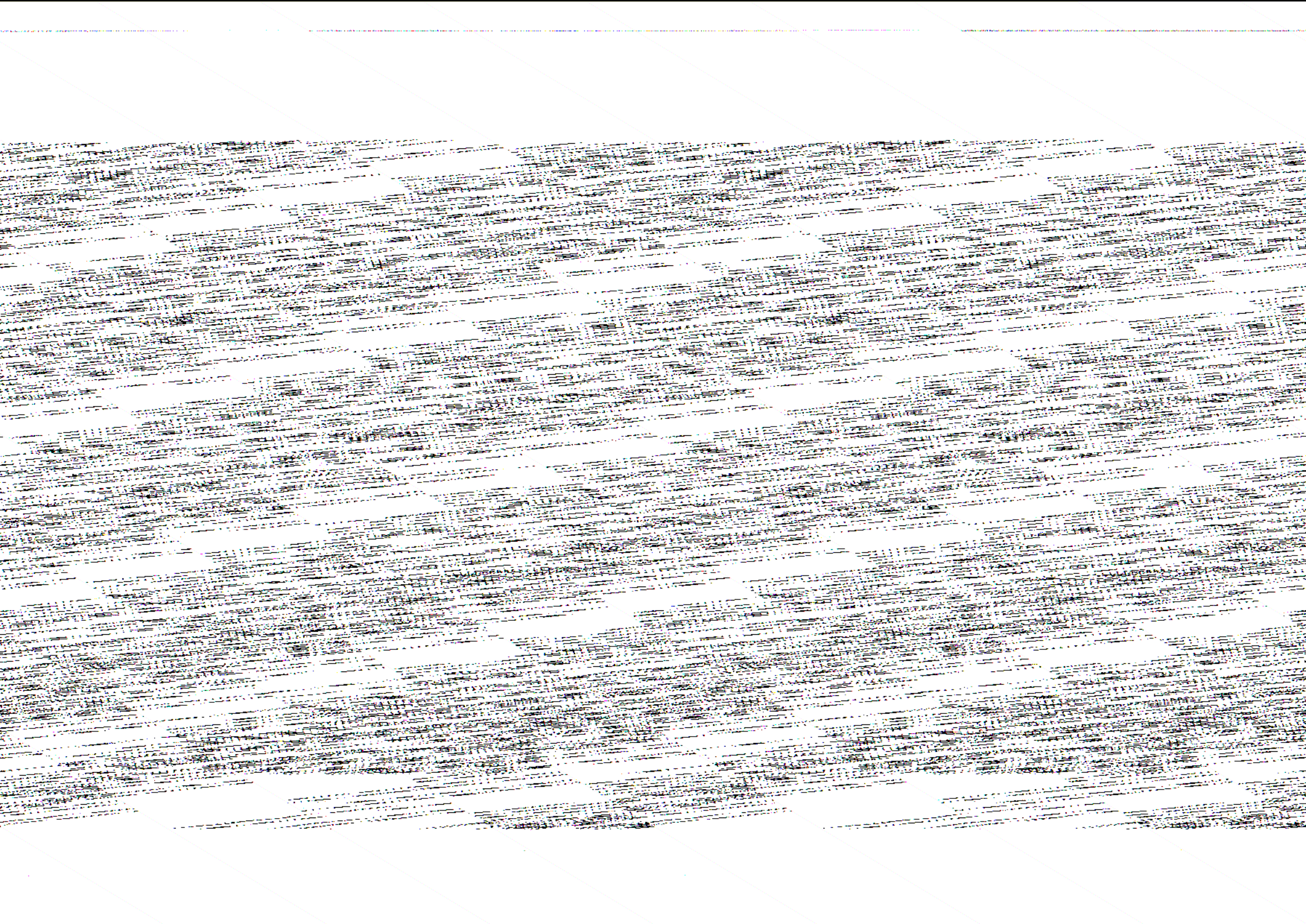


(c) As (a), but for a sampling at 1.0 Hz.



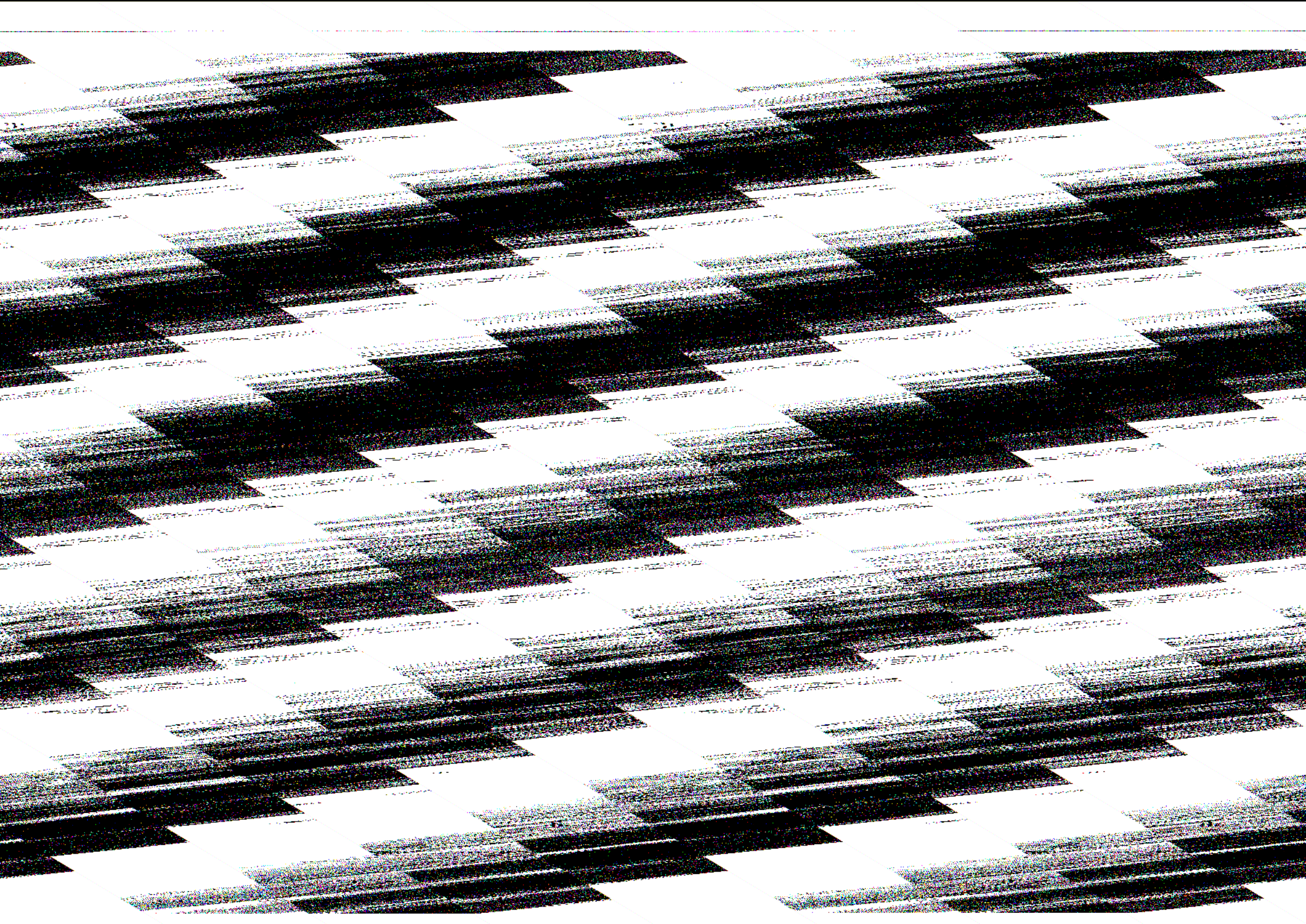
The pulse duration results in a range gate distance of ~272 m measured in a direction normal to the Earth's surface. We have computed equidistant vertical height levels separated by the range gate distance up to a height of 20 km, and, using a reference atmosphere, assumed that the DWL measures on the equivalent pressure levels. This will lead to a more or less random excess distance from the lowest range gate to the Earth's surface (geoid), rather than a fixed excess distance in case of fixed height levels. In case of the instrument scenarios I and II, only one in three vertical levels are kept, in which case the resulting resolution, resembles the forecast model used for the "nature" run (see Table 2.1).

The simulation of the DWL data distribution is discussed in section 4.1 and the simulation of HLOS winds at this distribution of points in section 4.3.3.



aspects of the ECMWF cloud scheme by comparison to visible and infrared images, an example of which is shown in Figures 3.1 and 3.2. Realistic cloud distributions are difficult to predict, but Figures 3.1 and 3.2 show that the synoptic representation of the ECMWF clouds is quite realistic. *Rizzi* (1994) compared TOVS raw radiances with simulated radiances over the full globe, implicitly verifying the amounts of cloud and cloud water/ice in the ECMWF model. Again a good synoptic description of the cloud amount can be found. However, the above studies also revealed some weaknesses. Although the amount of high level clouds (cirrus) is sufficient, the optical thickness (cloud water/ice content) is underestimated. The cloud amount at medium levels in the troposphere is underestimated, as is the cloud amount over sea behind cold fronts and in anticyclonic areas (low clouds). It is beyond the scope of this project to quantify (and correct for) these weaknesses.

Useful validation information is also available by comparison with a new prognostic cloud scheme which is being developed at ECMWF. It includes cloud water and ice as prognostic variables during the forecast. It is shown to be more realistic in, for instance, the prediction of the thickness of high level clouds, which may have consequences for the simulation of DWL observations; the lidar laser beam would generally penetrate less far into the troposphere and therefore give less information on the tropospheric wind. The cloud scheme used is for this reason probably a bit too optimistic in the amount of information acquired on the tropospheric wind, in cloudy areas in the tropics. On the other hand, high clouds cover a somewhat larger area in the current model than they do in reality, and in this sense the simulation is too pessimistic, over-estimating the areas in which the DWL sounding is blocked by cloud.



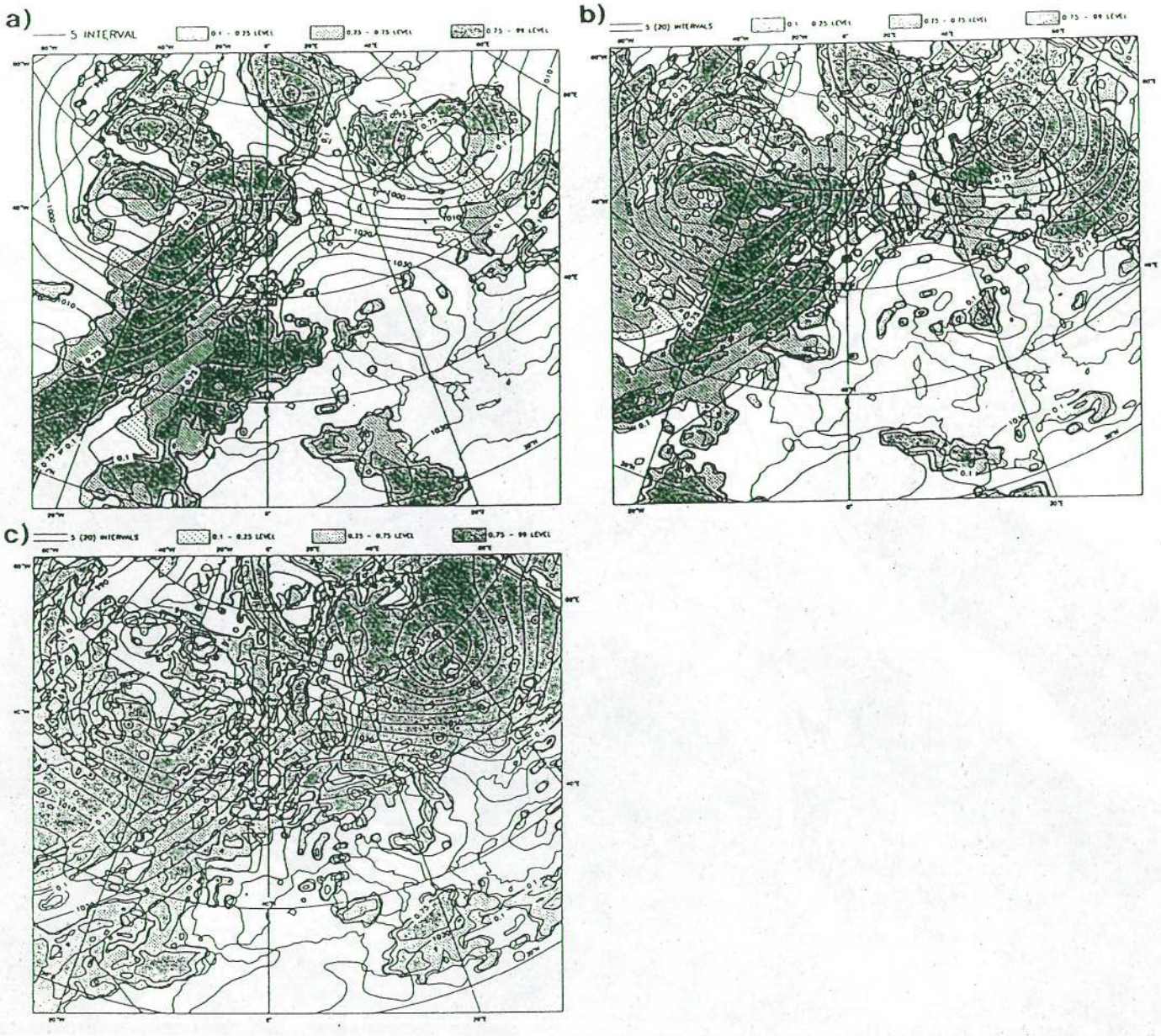


Fig 3.2:(a) Operational 36-hour forecast of high cloud cover, verifying at 15/1/93 00:00 UT over Europe. Contour levels are 10%, 25% and 75%. The cloudiness field overlays the corresponding mean sea level pressure forecast.

(b) As (a), but with medium cloud cover.

(c) As (a), but with low cloud cover.

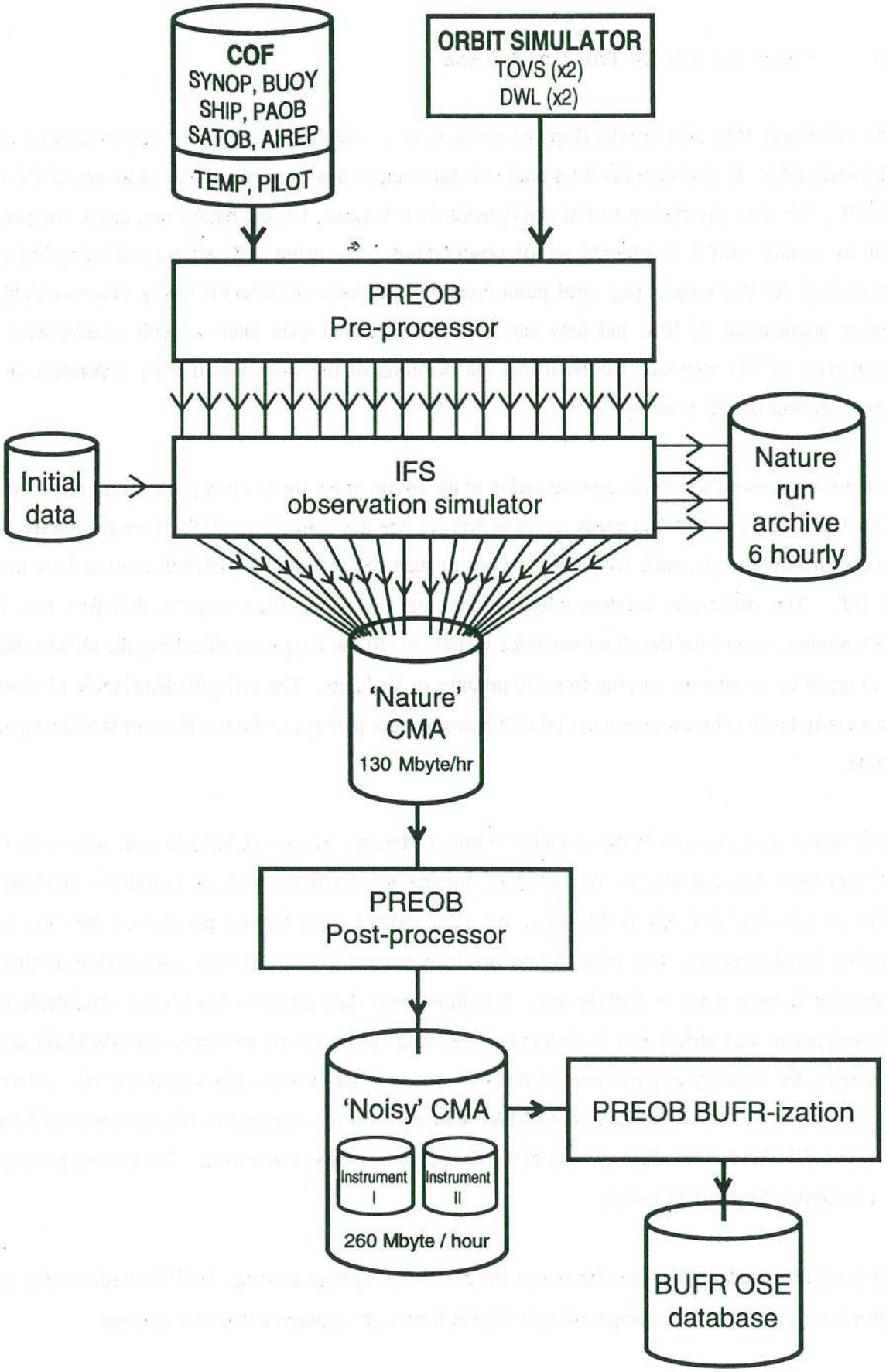
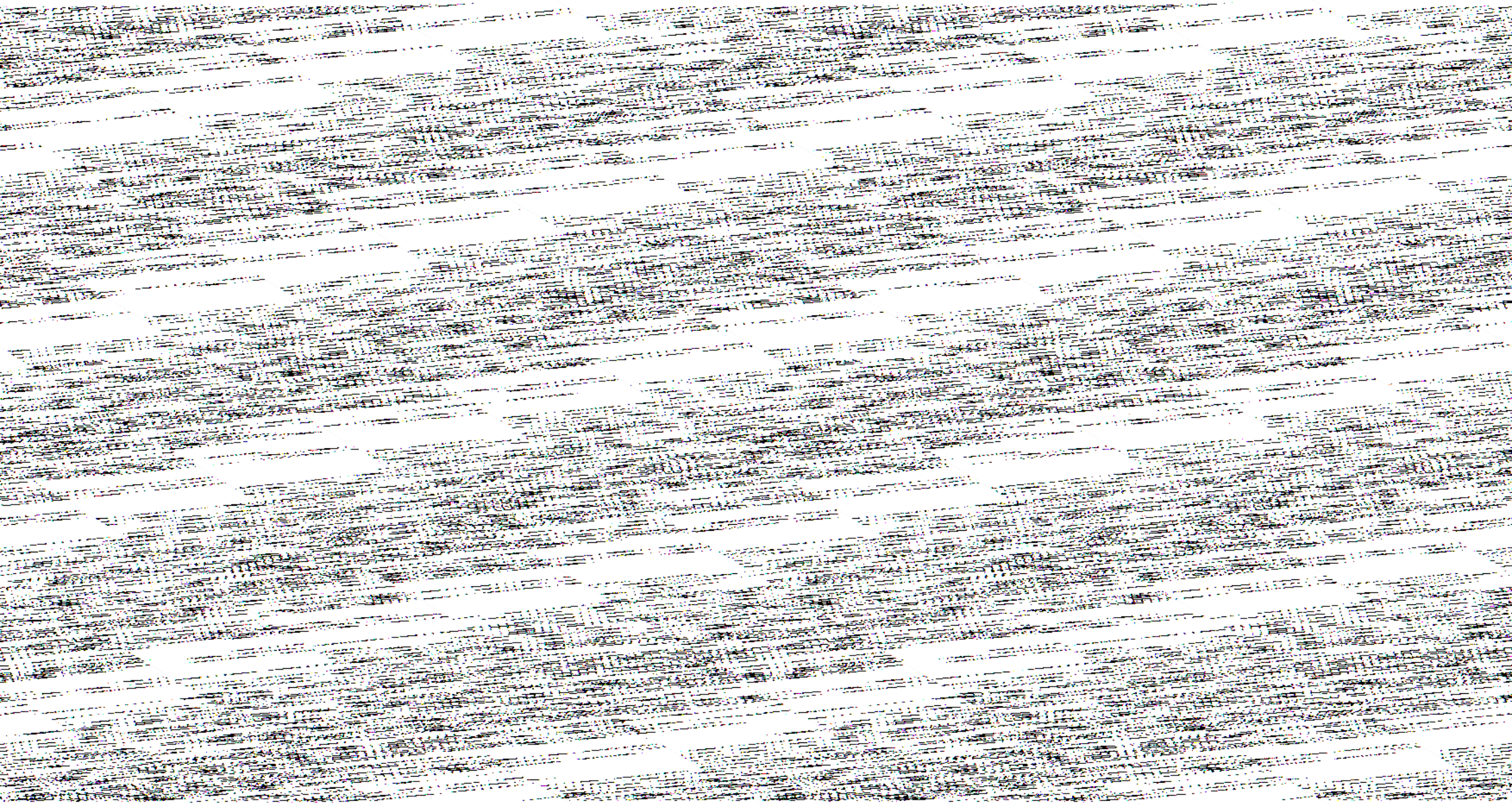


Fig 4.1: Schematic of the observation simulation system.



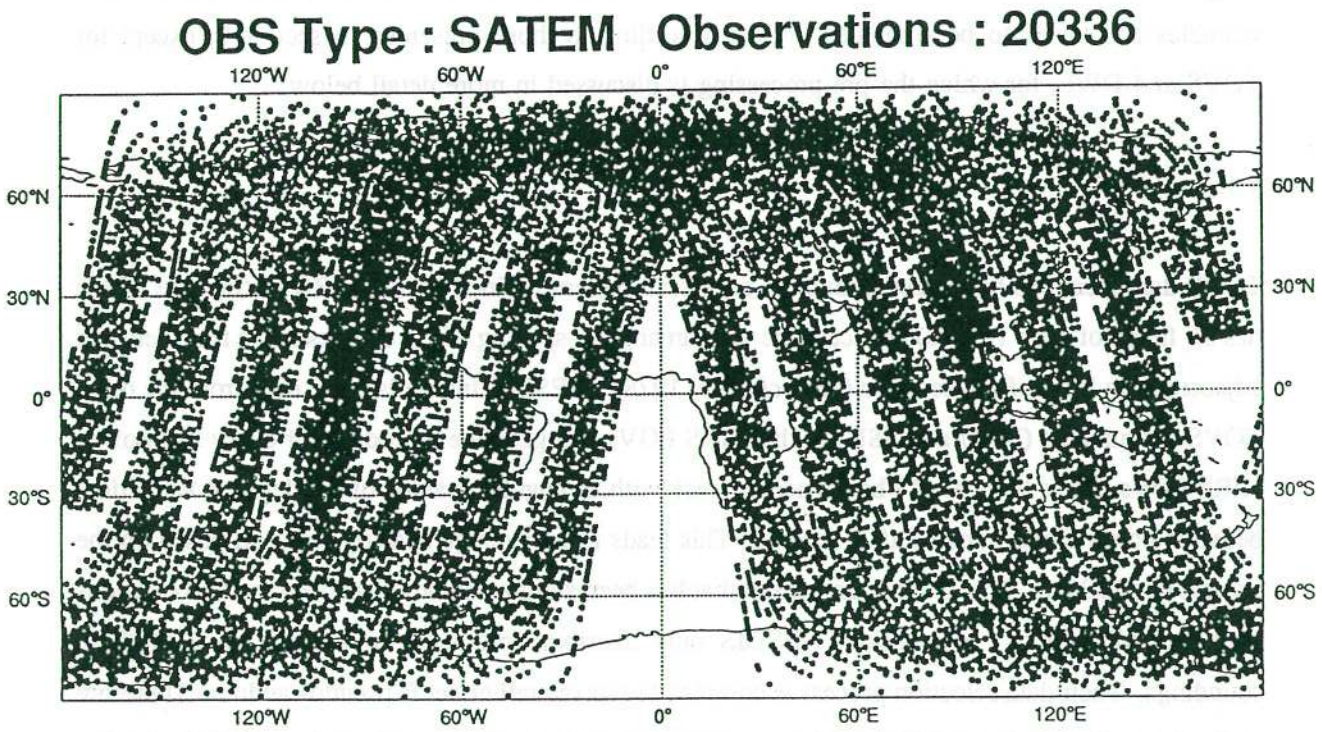


Fig 4.2: Simulated distribution of TOVS observations for two satellites from 09:00 UT until 15:00 UT 5/2/93. (Figure 2.2 gives a real distribution for NOAA-12 and NOAA-13 for a 6-hour period.)

For the simulation of TOVS data, the temperature and humidity have been interpolated on to the fixed pressure levels of the fast radiative transfer scheme RTTOV (*Eyre*, 1991), which acts as the "observation operator" for TOVS clear-column brightness temperatures. Above the top of the model (10 hPa) the temperature profile has been extrapolated using a climatological regression scheme. The nature run values of surface air and surface skin temperature have also been used as input to the radiative transfer calculation. The brightness temperatures have been calculated as if observed at nadir and with a surface emissivity of one, in order to imitate the effects of the pre-processing performed by NESDIS.

In order to store suitable cloud quantities with the simulated observations, the following developments were necessary. Prior to this project upper air forecast fields were not stored on a grid representation, but as a spectrum. The cloud variables, however, are computed in grid point space and possess large spatial variability. It was found that, through transformation to spectral space and back, substantial sub-scale noise was generated (*Rizzi*, 1994). Therefore a mechanism was developed to archive cloud-related fields on a grid representation, which was subsequently used to archive cloud cover and cloud water/ice.

The nature run is archived as follows. The cloud variables are stored with the other standard archive variables for a forecast experiment at 06, 12, 18 and 24 UT (end of day) for each one day run. In addition all the fields are stored for 00 UT on the first day of the experiment (except the cloud variables, which are all zero at this time). These fields constitute the nature run. They are archived in a compressed binary data format (GRIB), and they are available through standard field data retrieval tools. These archived fields would be suitable for comparison with similar fields generated by an OSSE on this database, or just to evaluate the data base. In addition, at the end of each day, a full dump is made of the meteorological fields available on the core memory, called the "restart file", to act as a starting point for the next day of simulation. Given the chaotic behaviour of the atmosphere and atmospheric models, this is the only way to ensure that 30 one-day forecasts result in the same state as one 30-day forecast, and it prevents so-called spin-up effects in the first few hours of each day, especially relevant for the cloud variables. These restart files would also be used when adding new observation systems to the data base.

4.3 Post-processor

The post-processor takes all simulated "true" observations and adds errors to them, according to the typical error characteristics of each data type.

4.3.1.1 Simulated error statistics for surface observations

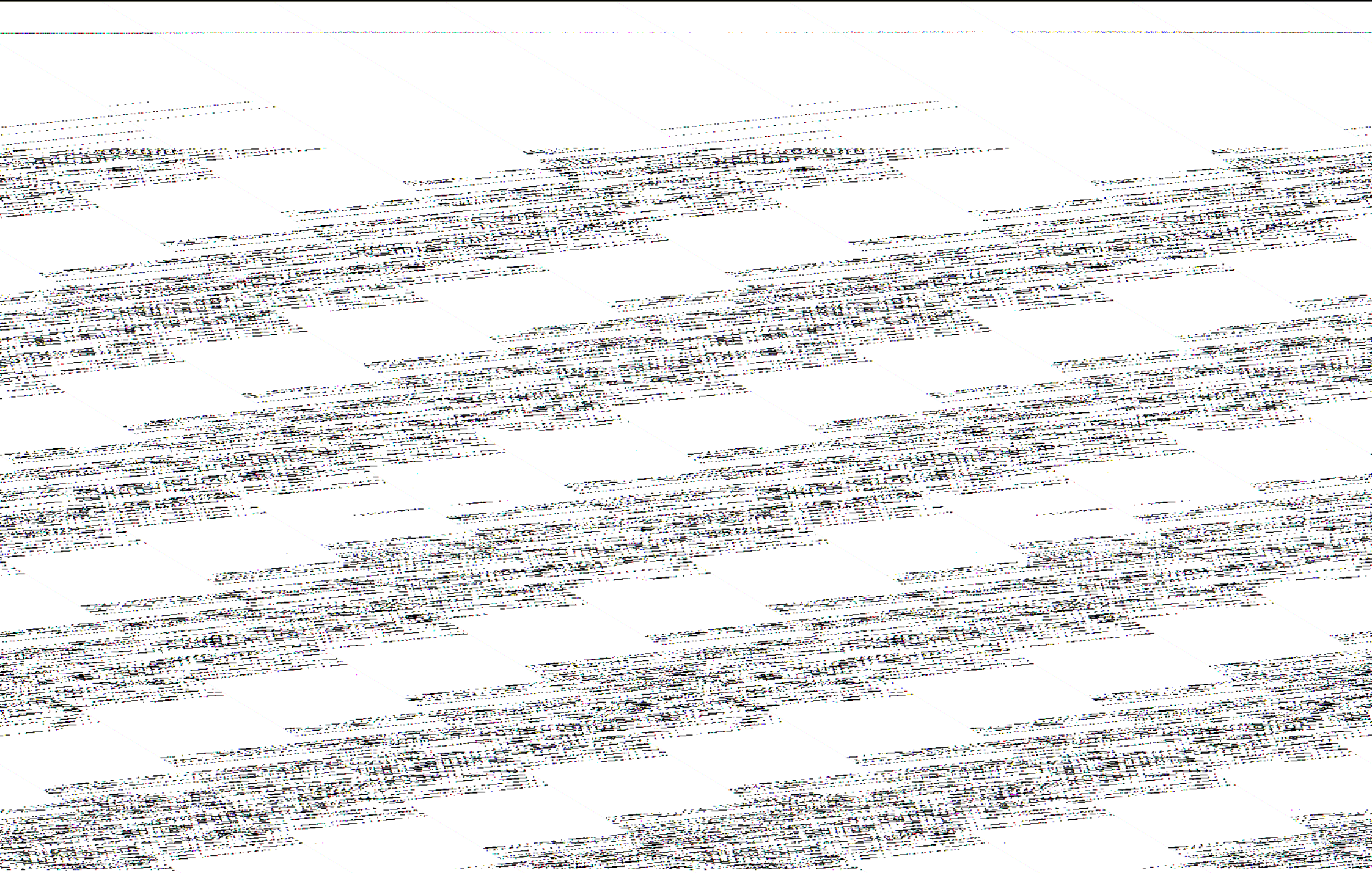
Table 4.1 shows the standard deviation of error, gross error rate and range of values for a gross error for simulated surface observations. Gross errors for wind are applied on the vector wind rather than on the individual wind components. The statistical values that are applied are stored in the CMA and BUFR files (see section 4.3).

Table 4.1: Standard deviation of error (a), gross error probability (b), and range of values for a gross error (c) for simulated surface observations. Empty table entries indicate combinations of observation type and variable that are not simulated (and do not exist).

	P_{msl} hPa	U,V m/s	T K	RH %
a) σ				
land SYNOP	1.0	2.0	2.5	13
SHIP	1.0	2.0	2.0	13
BUOY	1.3	2.2	2.0	13
PAOB	2.0			
b) P_G				
land SYNOP	0.015	0.020	0.020	0.040
SHIP	0.060	0.060	0.070	0.050
BUOY	0.030	0.040	0.030	0.040
PAOB	0.004			
c) R				
land SYNOP	31.25	23.6	33.3	100.0
SHIP	31.25	23.6	33.3	100.0
BUOY	23.26	70.7	100.0	100.0
PAOB	23.26			

4.3.1.2 Simulated error statistics for pressure level observations

From operational experience it is known that roughly 90% of all TEMPs contain no gross error in temperature or wind (*Gandin et al*, 1993). In Table 4.2 further figures are given for the number of TEMP reports with at least 1, 2, 3 and 4 gross errors in temperature or wind as monitored by the UK Meteorological Office (courtesy Richard Graham). Based on this information we decided to simulate 90% of TEMPs without a gross error, and the number of soundings with one or more gross errors simulated in the remaining 10% as shown in Table 4.2. In Table 4.3 the gross error rates apply for only those 10%.



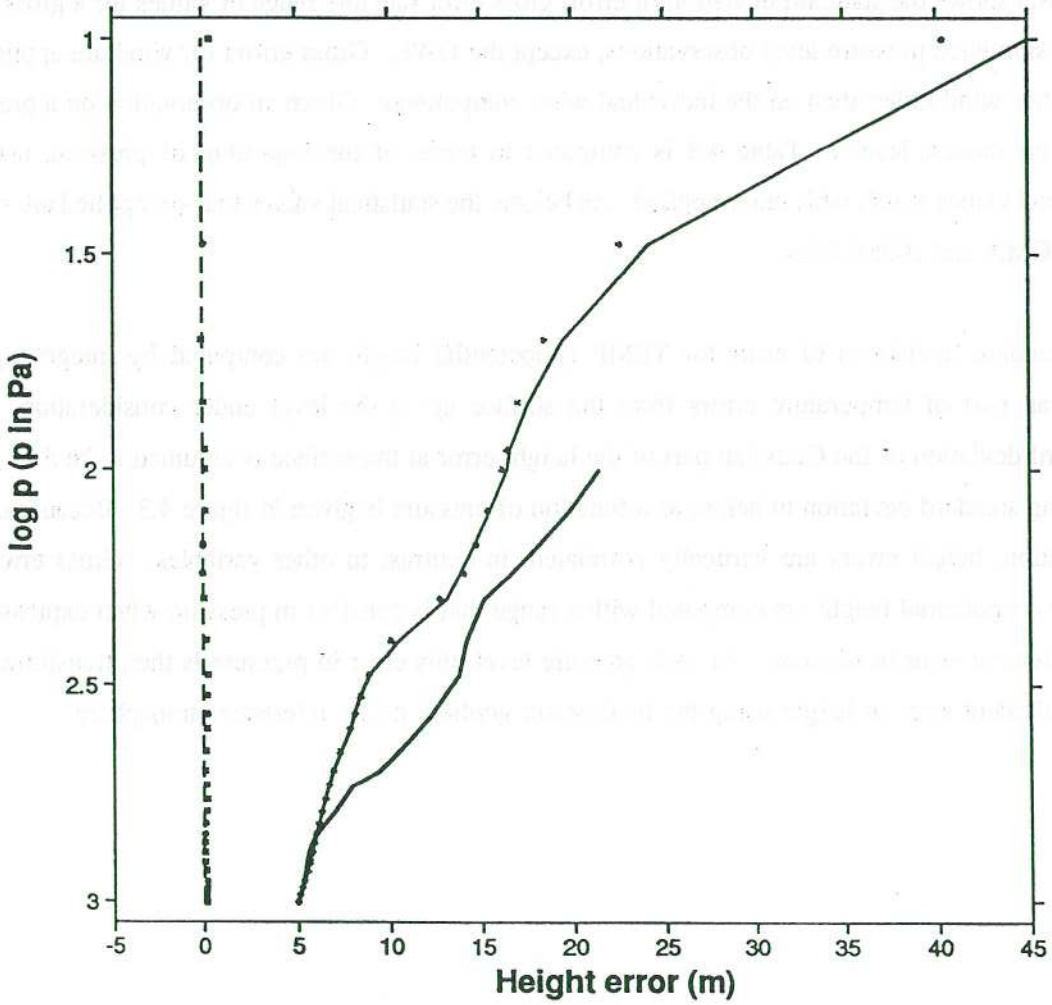
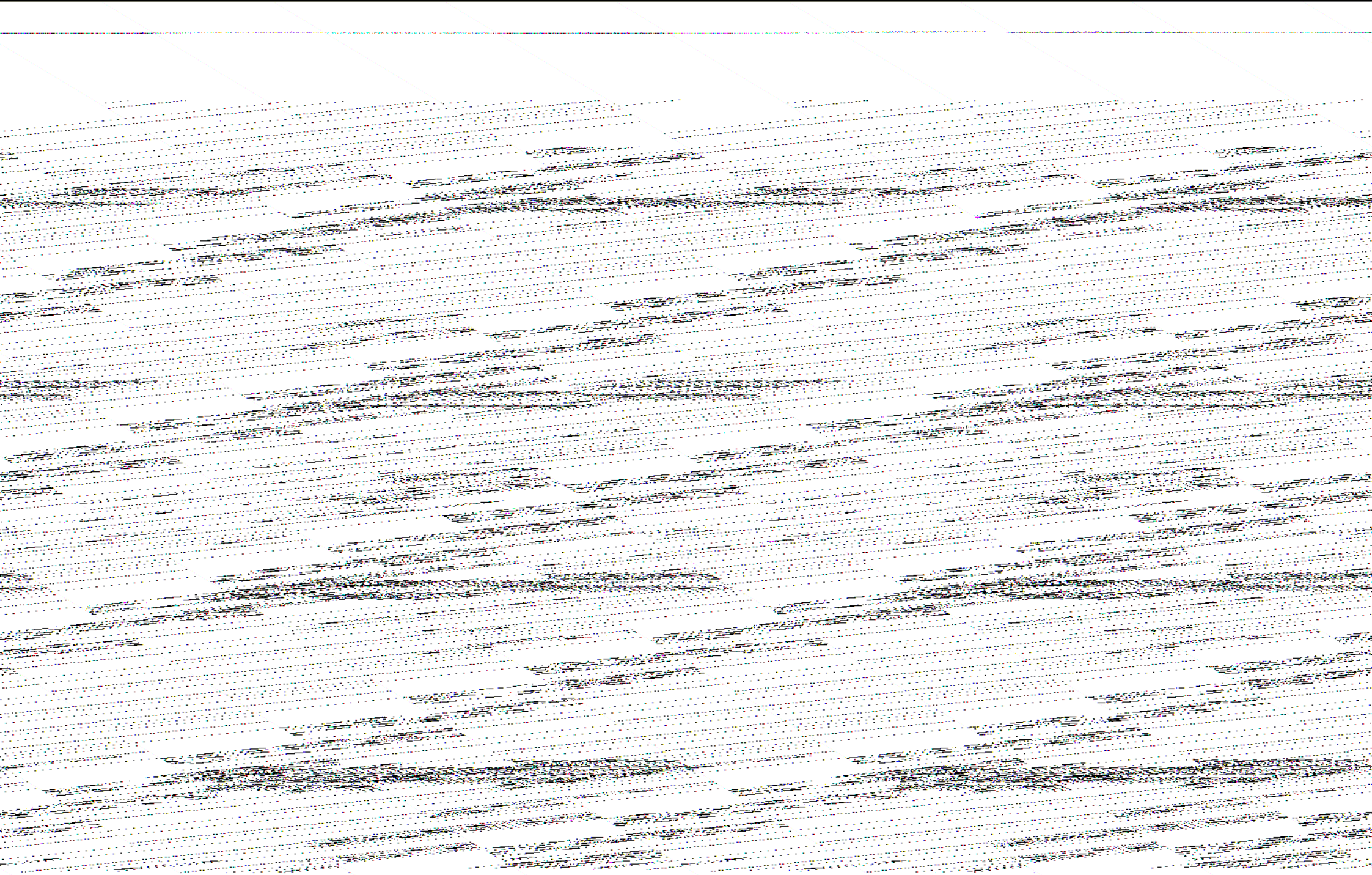


Fig 4.3: Simulated height bias and standard deviation of error as a function of height (symbols) after integration of 1000 profiles with Gaussian temperature errors as specified in Table 4.3. For comparison the theoretically expected bias and standard deviation of error are given by respectively the dashed and the solid line. The thick solid line represents the estimated standard deviation of error as used from the ECMWF OI analysis.



d)

	Range of gross error:			
	Z [hPa]	U,V [m,s]	T [K]	RH [%]
TEMPSs/PILOTs	23.25	70.71	100.0	100.0
AIREPs		70.71	33.3	
SATOBs		25.82		

4.3.2 *TOVS*

The first part of the post-processing is the choice of the cloud-clearing route based on the total fractional cloud cover, N. It should be noted that N is the fractional cloud cover on the scale represented by the nature run (~100 km), which is much larger than the HIRS FOV (~20 km). Thus each HIRS FOV does not "see" a fractional cover N, but rather one that is related to it statistically. In practice, a FOV is found to be "clear" if the fractional cloud cover within the HIRS FOV is close to zero, and "cloudy" if it is close to one. For intermediate values it can be either "clear", "partly cloudy" or "cloudy" depending on the nature of the cloud. We have simulated these effects using the following algorithm to calculate the probability of cloud-clearing route p_i given N:

For the "clear" route, p_1

$P(p_1|N) = 1$, for $N < a - d/2$

$= 0.5 - (N-a)/d$, for $a - d/2 < N < a + d/2$

$= 0$, for $N > a + d/2$.

For the "partly cloudy" route, p_2

$P(p_2|N) = 0$, for $N < a - d/2$

$= 0.5 + (N-a)/d$, for $a - d/2 < N < a + d/2$

$= 1$, for $a + d/2 < N < b - d/2$

$= 0.5 - (N-b)/d$, for $b - d/2 < N < b + d/2$

$= 0$, for $N > b + d/2$.

For the "cloudy" route, p_3

$P(p_3|N) = 0$, for $N < b - d/2$

$= 0.5 + (N-b)/d$, for $b - d/2 < N < b + d/2$

$= 1$, for $N > b + d/2$.

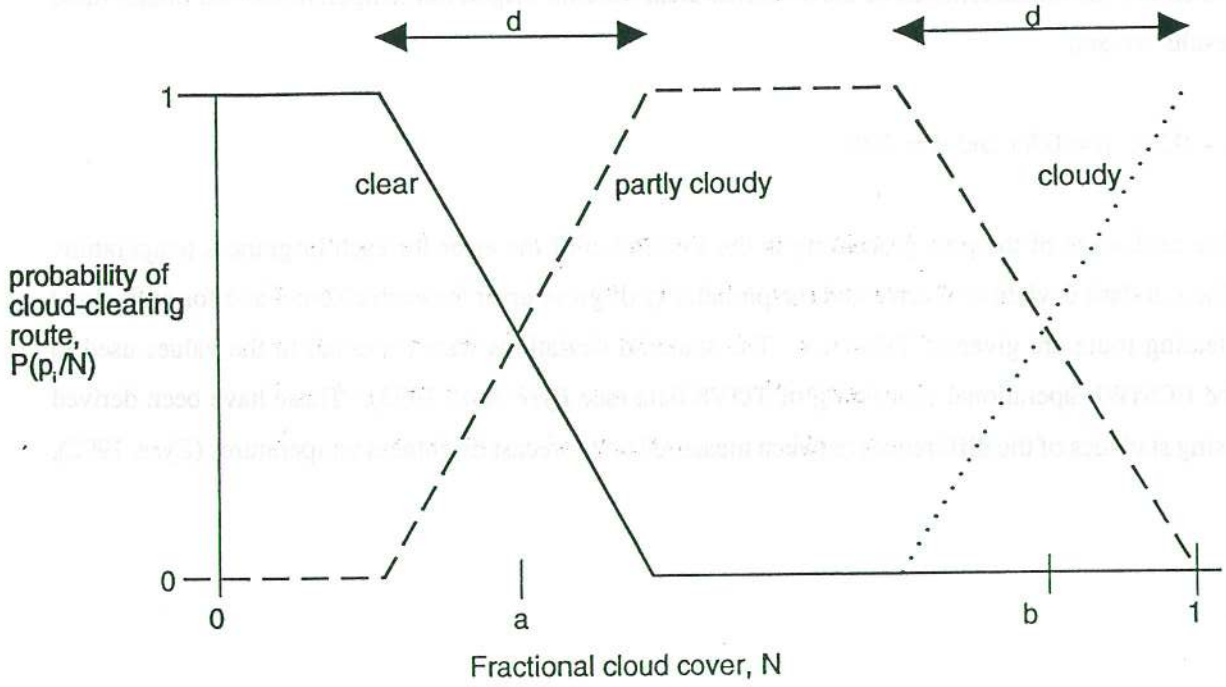


Fig 4.4 Illustrating the cloud-clearing route decision algorithm used for simulated TOVS data.

4.3.3 *DWL*

The errors we provide for the DWL constitute an instrument and processing (detection) error depending on instrument and atmospheric optical properties, and a representativeness error. This latter error is a measure of how the spatial and temporal scales of a particular measurement represent the spatial and temporal scales resolved by the equivalent NWP model variable. We first discuss the detection error derivation.

The signal bandwidth during processing determines the accuracy of LOS wind speed detection. A smaller bandwidth results in better accuracy. A priori information on LOS wind speed is therefore essential in the processing. Let us suppose V_b is our a priori wind speed estimate along LOS. Given the bandwidth B we can compute the range of detectable wind speeds, which is $[V_b-R/2,\ V_b+R/2]$, where R=0.5 B λ (91 m/s). For the case of argument we subtract V_b from the LOS wind and assume that the error in V_b is much smaller than R.

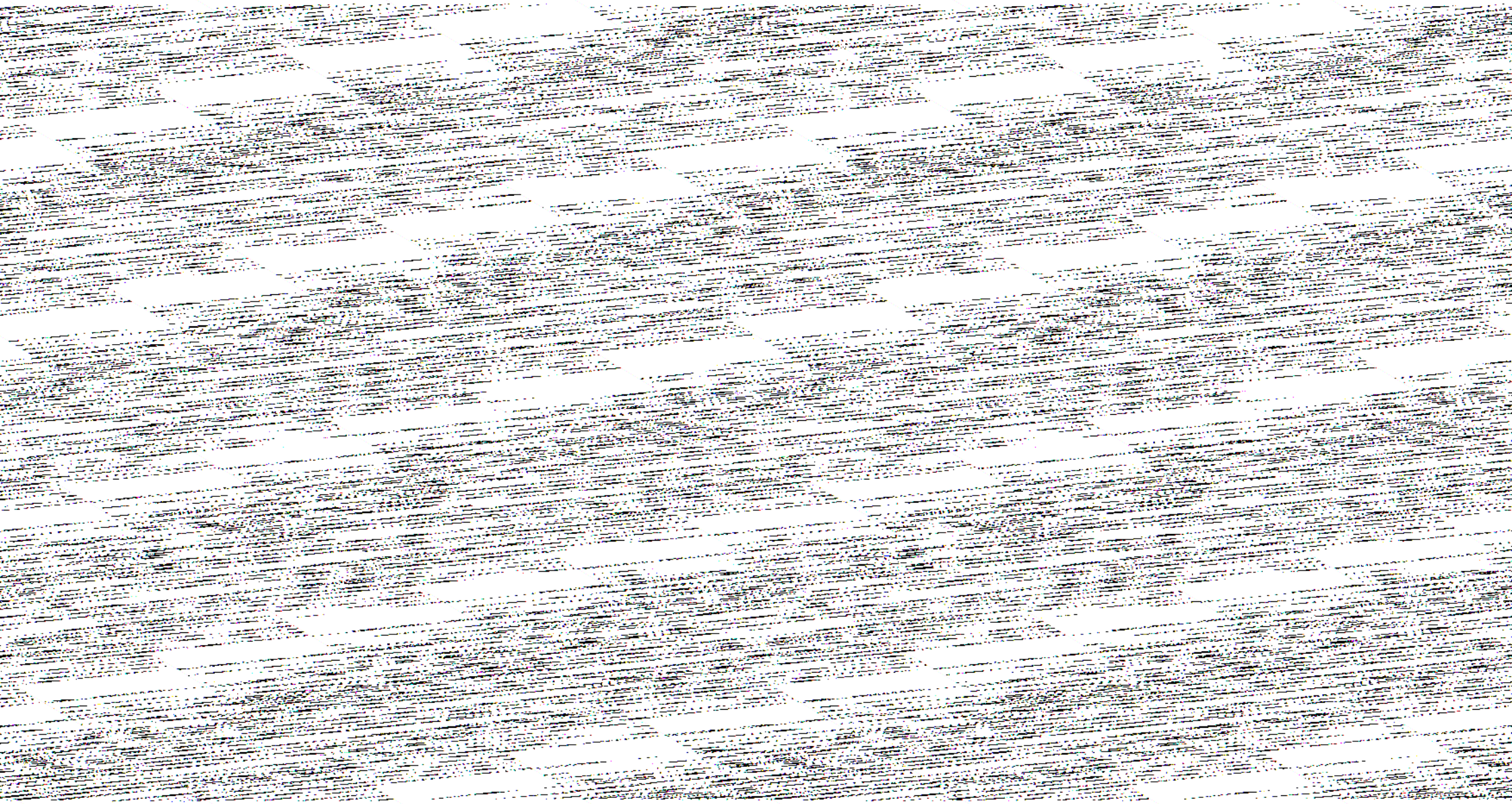
According to *Lorenc et al* (1991) we wish to describe the LOS speed probability over the range R as:

$$P(V) = (1-P_G)\ N(0.0,\sigma) + P_G/R \qquad (1)$$

Where $N(0.0,\sigma)$ describes a normal distribution with mean 0.0 and standard deviation σ, P_G is a gross error probability constant over the range $[-\frac{1}{2}R,\ \frac{1}{2}R]$, and σ is the standard deviation of error for speeds without a gross error. In *Courtier et al* (1992) the Zrnic equation is given as:

$$\sigma = \frac{\lambda}{2\sqrt{M}T_S}\sqrt{\frac{\sigma_D\ T_S}{4\ \sqrt{\pi}} + \frac{2\ \sigma_D^2\ T_S^2}{SNR} + \frac{1}{8\ \pi^2\ SNR^2}} \qquad (2)$$

with σ the standard deviation as a function of signal-to-noise ratio (SNR) for pulse-pair processing of DWL measurements. M is the number of samples used in the processing, T_S the sampling period, λ the laser wavelength, and σ_D the signal Doppler width. The equation is based on small perturbation theory and normal error statistics, and therefore not valid for low SNR. For $SNR < 0$ dB, the third term under the square root is the most important, and σ does not depend on σ_D.



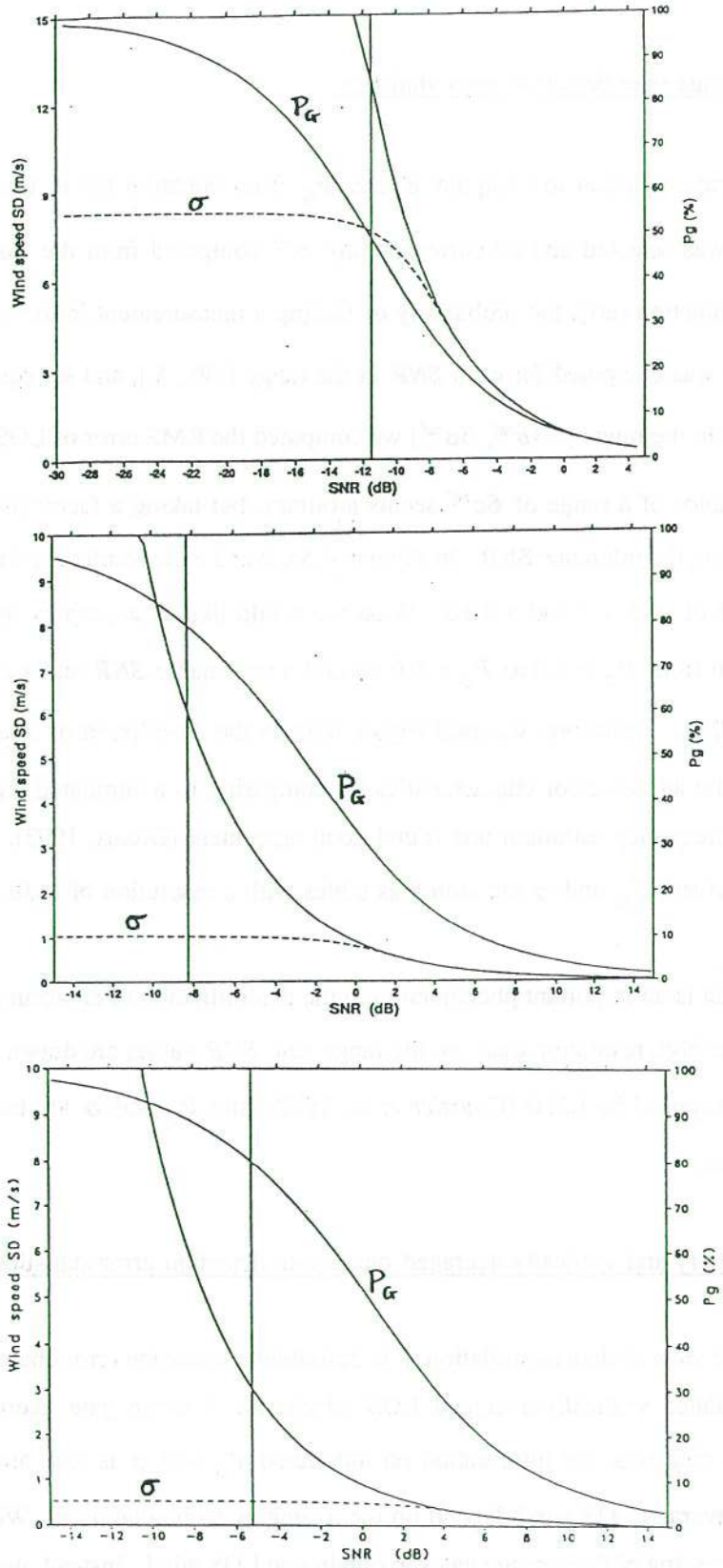
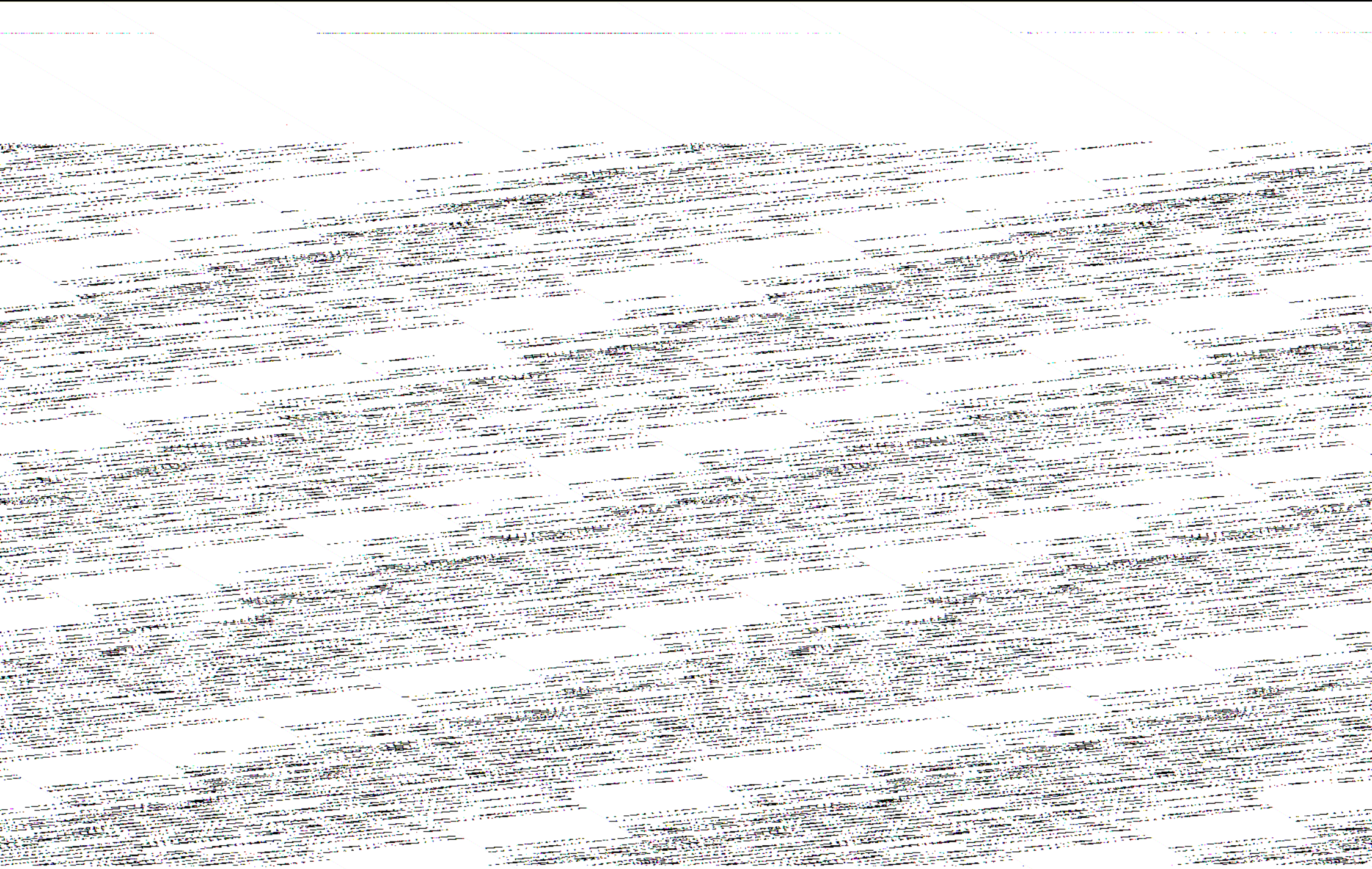


Fig 4.5:(a) Simulated values of LOS σ (dashed) and P_G (dotted) as a function of measured SNR for a "3-σ" test based on a reference value of σ, obtained from the Zrnic equation (solid - see text) and a reference SNR of -7.5 dB.

(b) As (a), but for a SNR of 1.5 dB.

(c) As (a), but for a SNR of 5.0 dB.



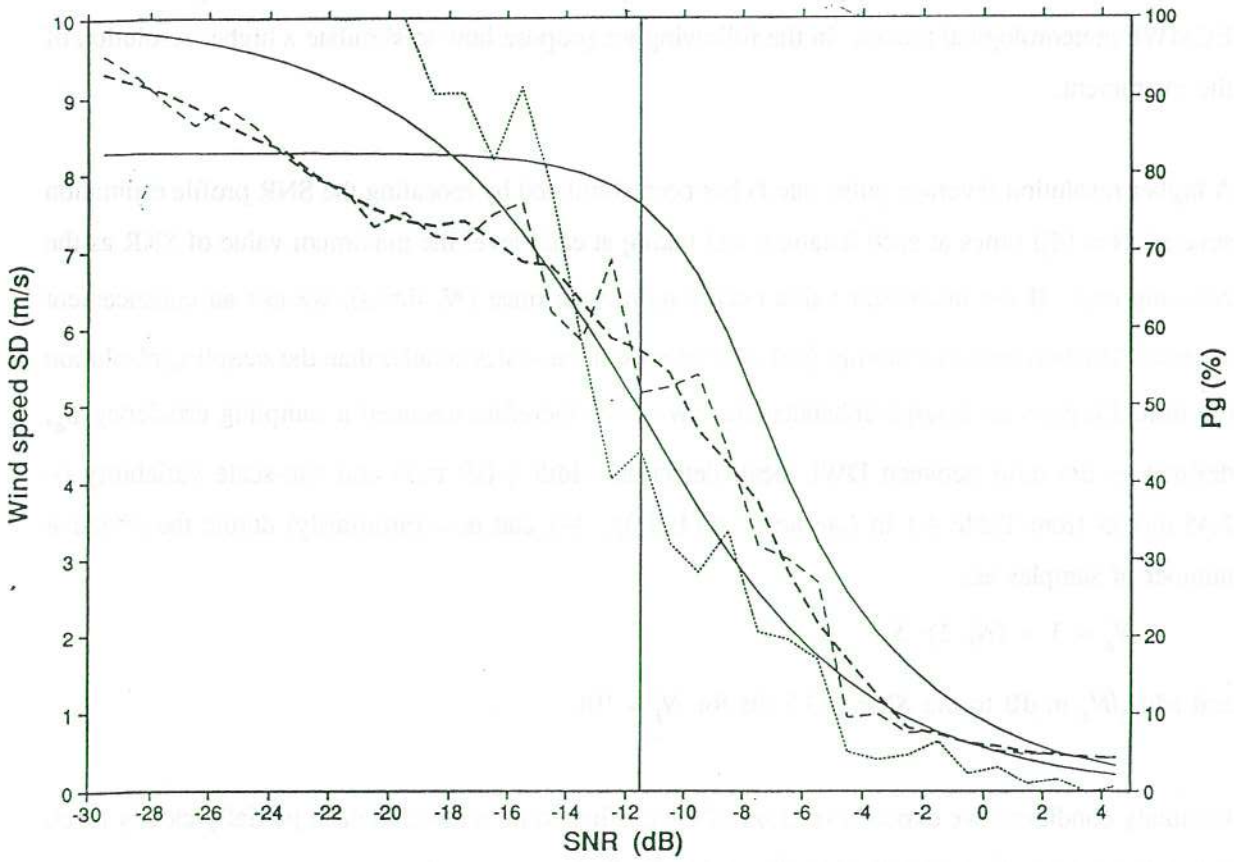
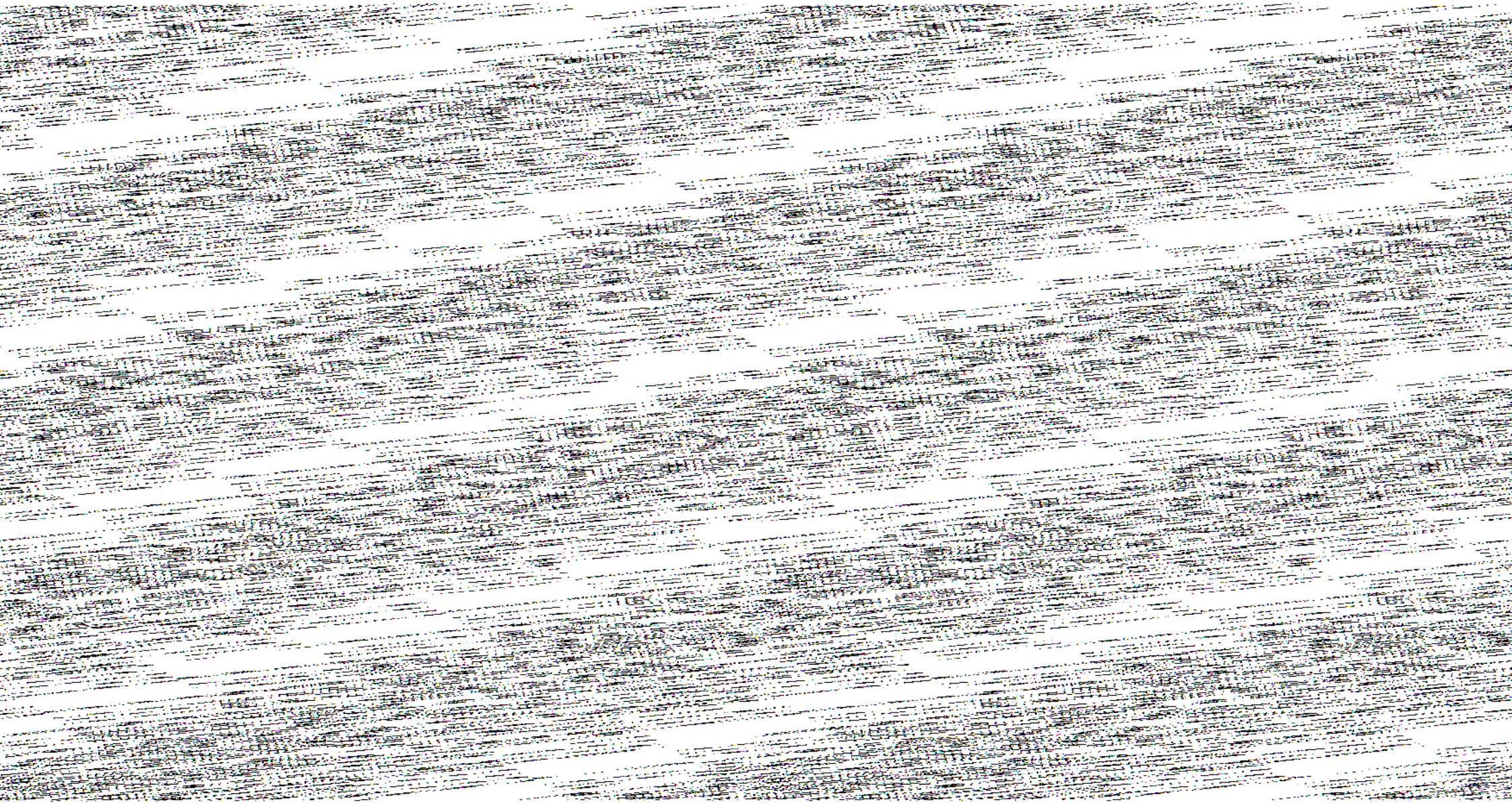


Fig 4.6: Simulated values of LOS σ and P_G as a function of processed SNR, representing a vertical average over three range gates. The thin solid line represents the σ curve given in Figure 4.4a. The thin dotted (P_G) and dashed (σ) line are the result of the procedure described in Appendix A. The thick dashed line is the σ obtained after using a 5-point smoother.



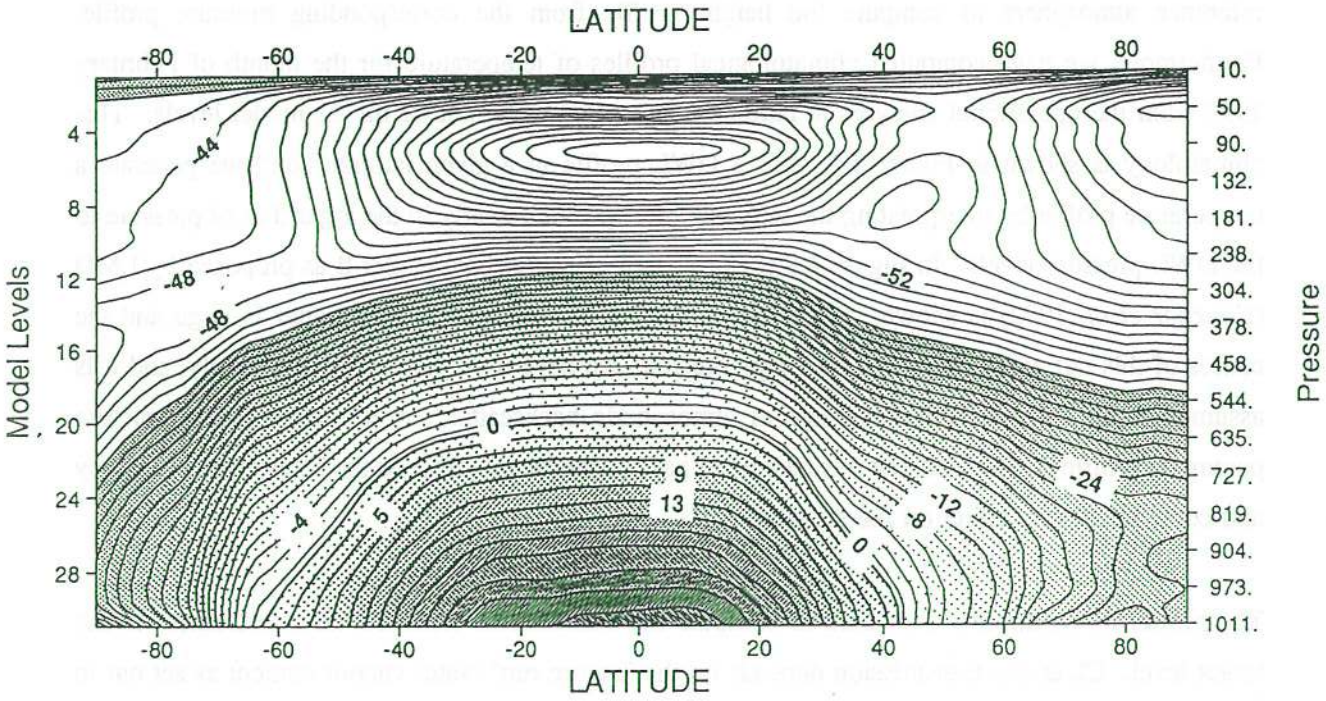


Fig 4.7: Mean zonal temperature as analyzed by the ECMWF model over the month of February 1993.

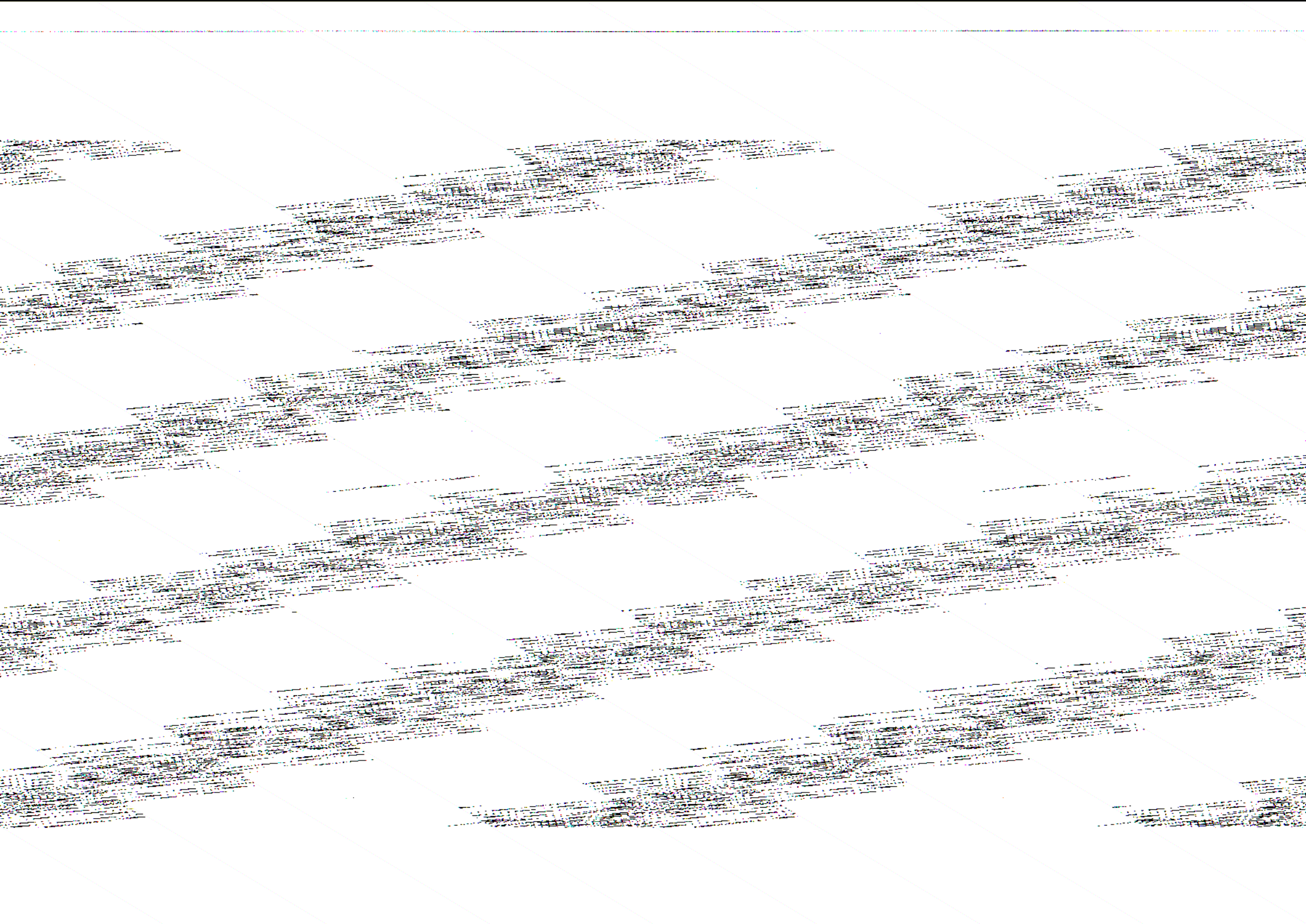
4.3.3.4 Addition of the representativeness error

The detection error and representativeness error are squared and added, and the square root is taken to give the final standard deviation of error for the HLOS wind. We assumed that the representativeness error for a 300 m and a 900 m average are the same, and as given in Table 4.5. A dependence on meteorological conditions (e.g. such as Clear Air Turbulence (CAT)) as suggested in *Lorenc et al* (1991) is not implemented.

In cases with large backscatter, DWL LOS measurements are anticipated to have an accuracy of 1 m/s or less. This accuracy is certainly adequate, and it is clear that the main uncertainty will then come from the representativeness error. The magnitude of the representativeness error is mainly determined by the lack of horizontal extension of the measurement volume, and is therefore comparable to the representativeness error of radiosonde wind data. In certain cases, however, it is expected that the effective vertical extent of the measurement volume will be decreased due to the presence of clouds or aerosol stratification. This will increase the representativeness error of DWL data with respect to radiosondes. Instrument II (see section 2.3) would generally only provide information under those conditions, and therefore suffer from increased representativeness errors.

Table 4.5: Estimated representativeness error for DWL LOS winds. The error at the 1000 mb level contains a 0.7 m/s RMS contribution from PBL turbulence. Table is derived from Table 4.4 in *Lorenc et al* (1991).

Pressure (hPa)	Representativeness σ (m/s)
100-10	1.8
150	1.7
200	1.7
250	1.7
300	2.4
400	2.7
500	2.8
700	2.4
850	2.2
1000	1.7



5. VALIDATION

The steps necessary to generate the data base need validation. The first step is the production of the nature run. In our case this was done simultaneously with the second step which is the simulation of "nature" observations. A third step is to disturb those observations with the known noise characteristics of each system, and this results in the actual data base. The validation of these three steps is discussed in sections 5.1, 5.2 and 5.3 respectively.

5.1 Nature run

The actual weather conditions from 5/2/93 to 7/3/93 were discussed in section 3. The weather regimes during a 30 day forecast starting from 5/2/93 are bound to be different due to the chaotic behaviour of the atmosphere, especially in the latter part of the period. The nature run is archived every 6 hours for a standard set of relevant dynamical and physical variables. The weather regimes during the 30-day forecast can easily be obtained from these archived fields.

A new set of archived variables consists of the 3-dimensional cloud amount and cloud water/ice grid point fields, which are also stored 6-hourly. Storing these variables on a 3D grid point field prevents the introduction of noise due to spectral truncation. We validated the 3D structure of the clouds and the cloud water/ice as archived, and found distributions as expected.

For the observation simulator, IFS is run in a mode where the nature run is created simultaneously. It is expected that the simulation of observations has no effect on the result of the forecast. To test this we ran a normal 2-day forecast from the same initial conditions as from which we started the observation simulator on 5/2/93 at 00 UT. We verified that the meteorological states of the observation simulator and the off-line forecast were identical on 6/2/93 00 UT and 7/2/93 00 UT down to CRAY machine precision. Following days use the same procedure to restart from the previous day, as was used for day 1 to day 2. Therefore a test with a 2-day forecast should suffice.

5.2 Observation simulator

Figure 5.1 shows a verification of the "nature" observations against a field of the nature run. Similar verifications were done for all simulated variables at levels around 1000 mb, 850 mb, 700 mb and 100 mb.

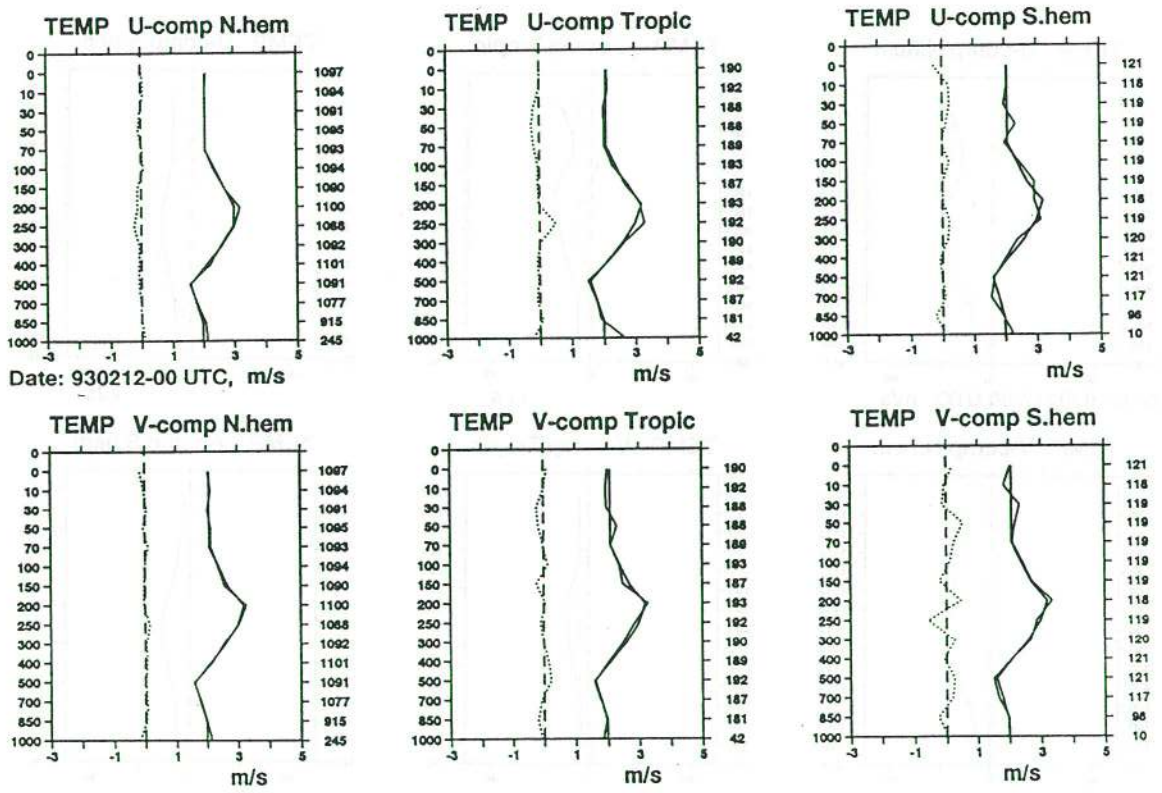
A second step was to verify that the mean and SD of the simulated observations for the different observation types, variables and geographical areas is as expected. Figure 5.2a shows such a verification. A third and very exacting test was to compute the observation equivalent in a second pass and subtract the obtained value from the "nature" observation value. Since the second pass is scientifically equivalent, but technically not the same, the test verifies the robustness of the software to compute observation equivalent values from the "nature" run.

5.3 Data base

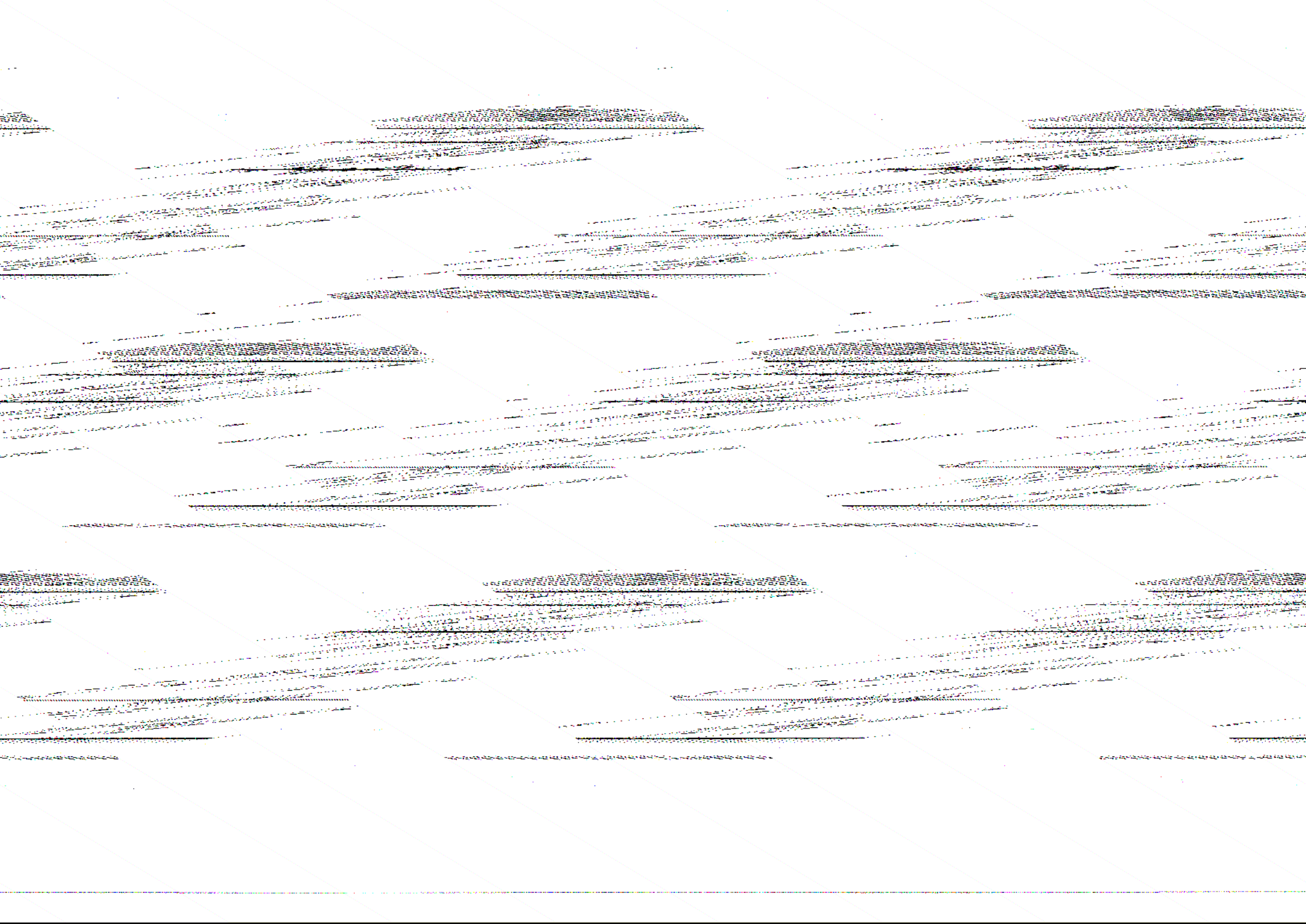
Obviously, after applying the post-processor to the "nature" observations, departure statistics of simulated observations minus "nature" run should reveal exactly the statistical properties applied. Figures 5.2b and c show such statistics separately for Gaussian and gross errors, and verification with the values presented in section 3 shows they are as expected.

The DWL post-processor is the most complicated and warrants more extensive testing. The DWL post-processor was therefore tested separately step by step with different test atmospheres, mainly varying in cloud amount and cloud thickness. Enhanced back-scattering from cloud and cloud transmission (cirrus) were found to be realistic, as well as the output values of SNR, σ and P_G. After these off-line tests the DWL post-processor was interfaced to process CMA files. Two examples of input and output profiles of temperature, humidity, cloud variables, backscatter, SNR, σ and P_G are shown in Figures 5.3 and 5.4.

The External organization of the database, and the ways of access to it, are described in Appendix E. A detailed user manual to read and decode the data can be found in Appendix F.



(b) As (a), except mean and standard deviation of observation errors (gross errors excluded).



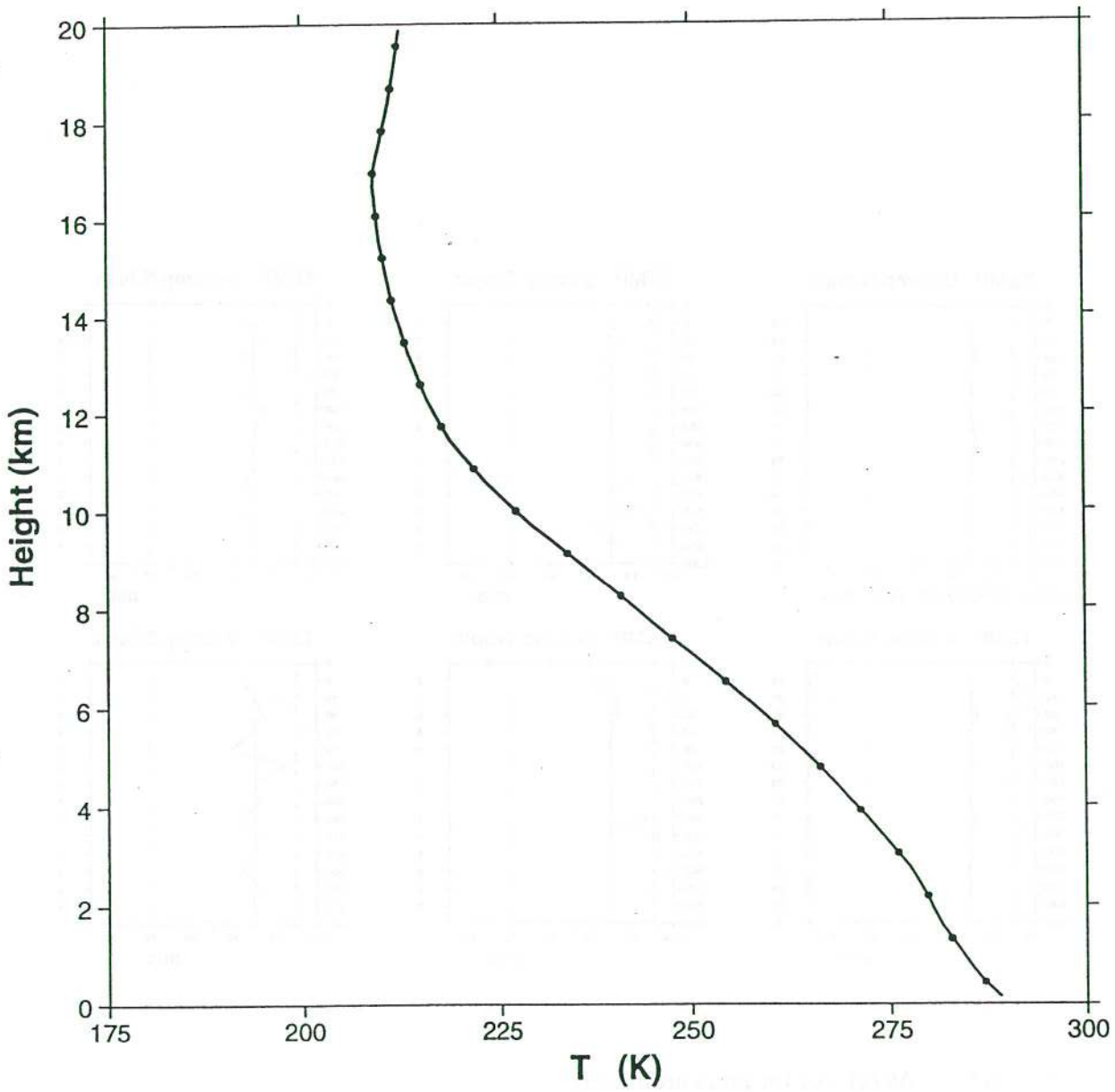
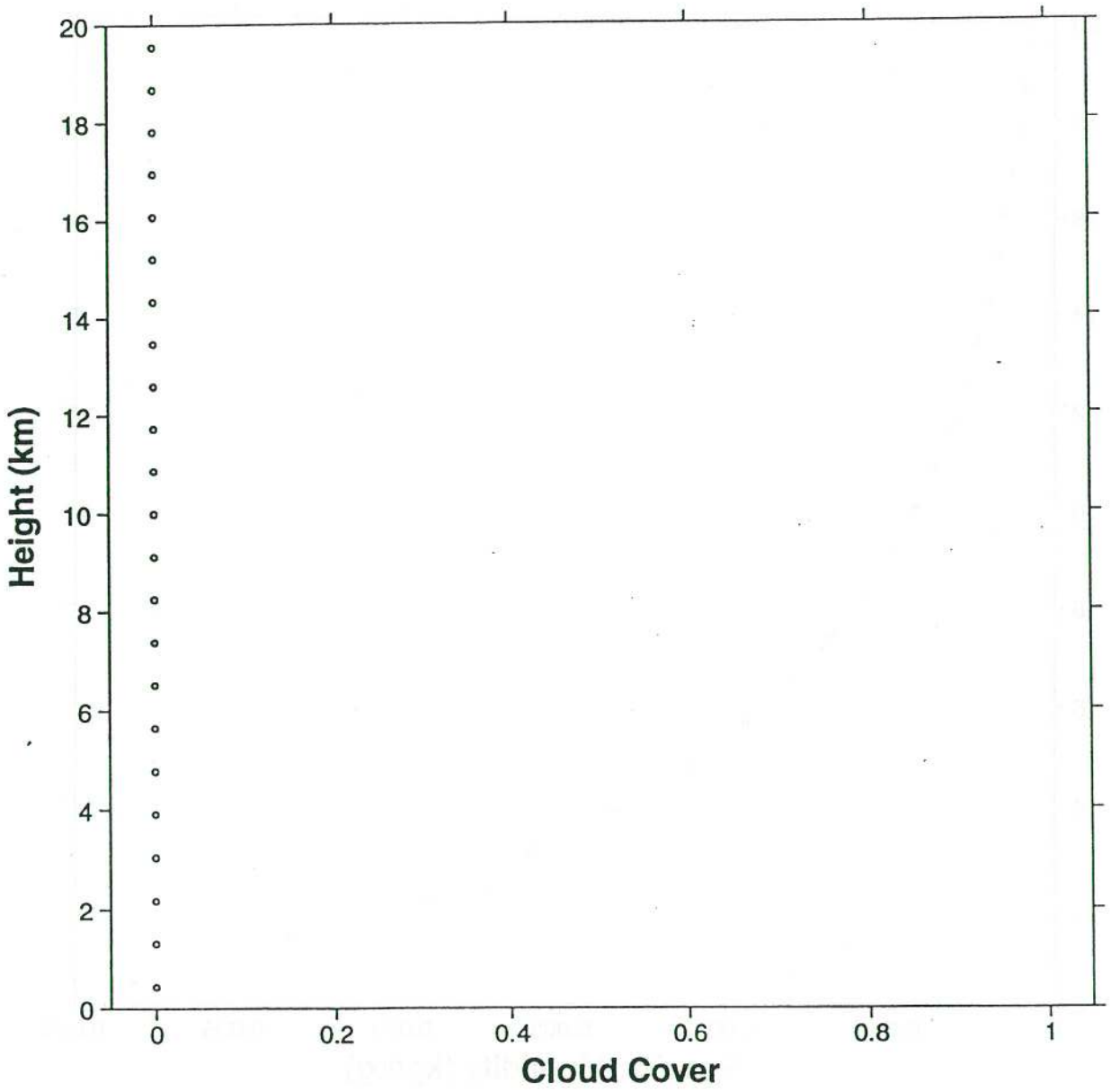
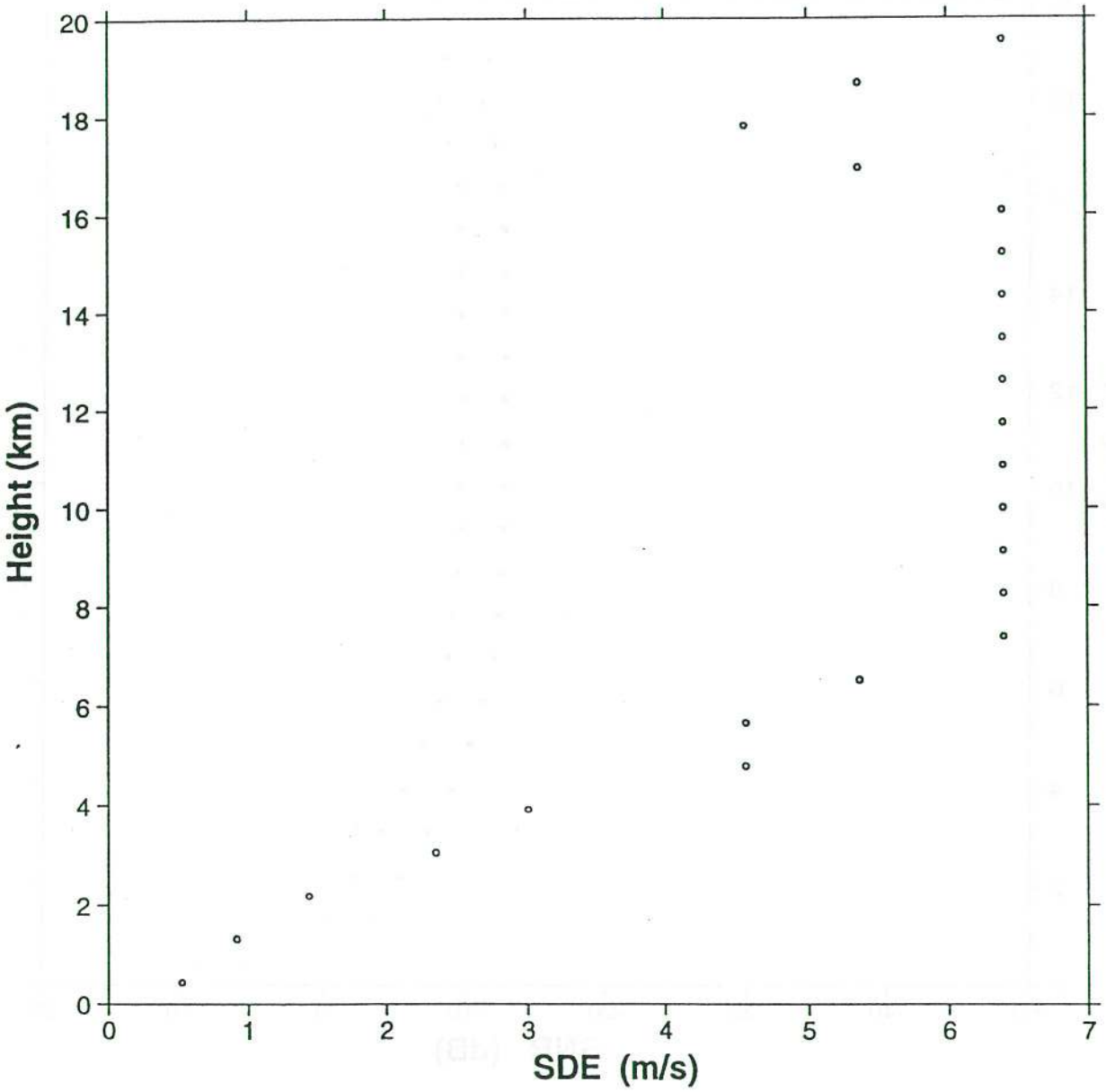


Fig 5.3: Profiles of variables used or computed in the DWL post-processor.
(a) Temperature profile.



(c) As (a), but for cloud cover.



(e) As (a), but for detection σ (excludes representativeness σ).

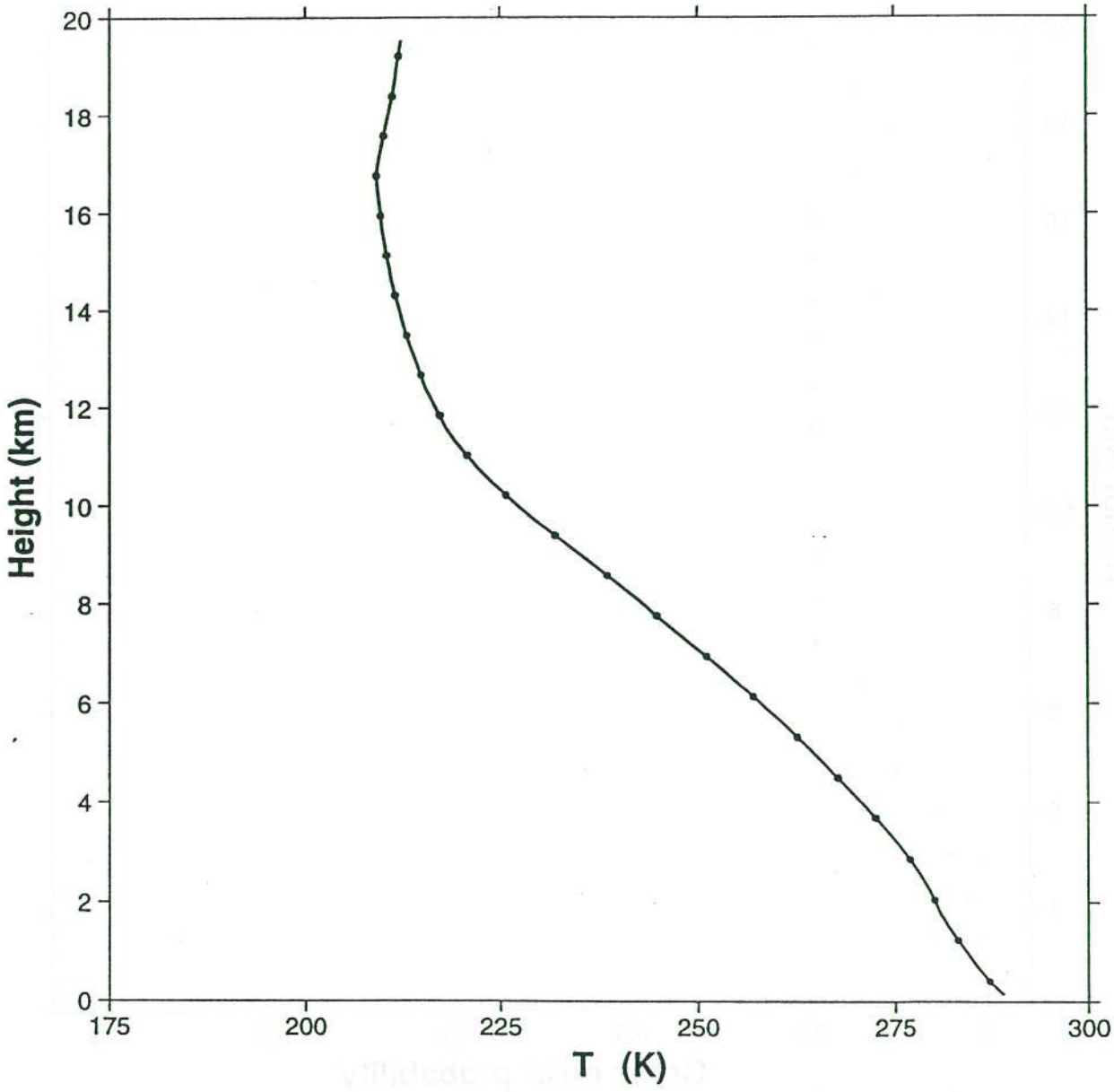
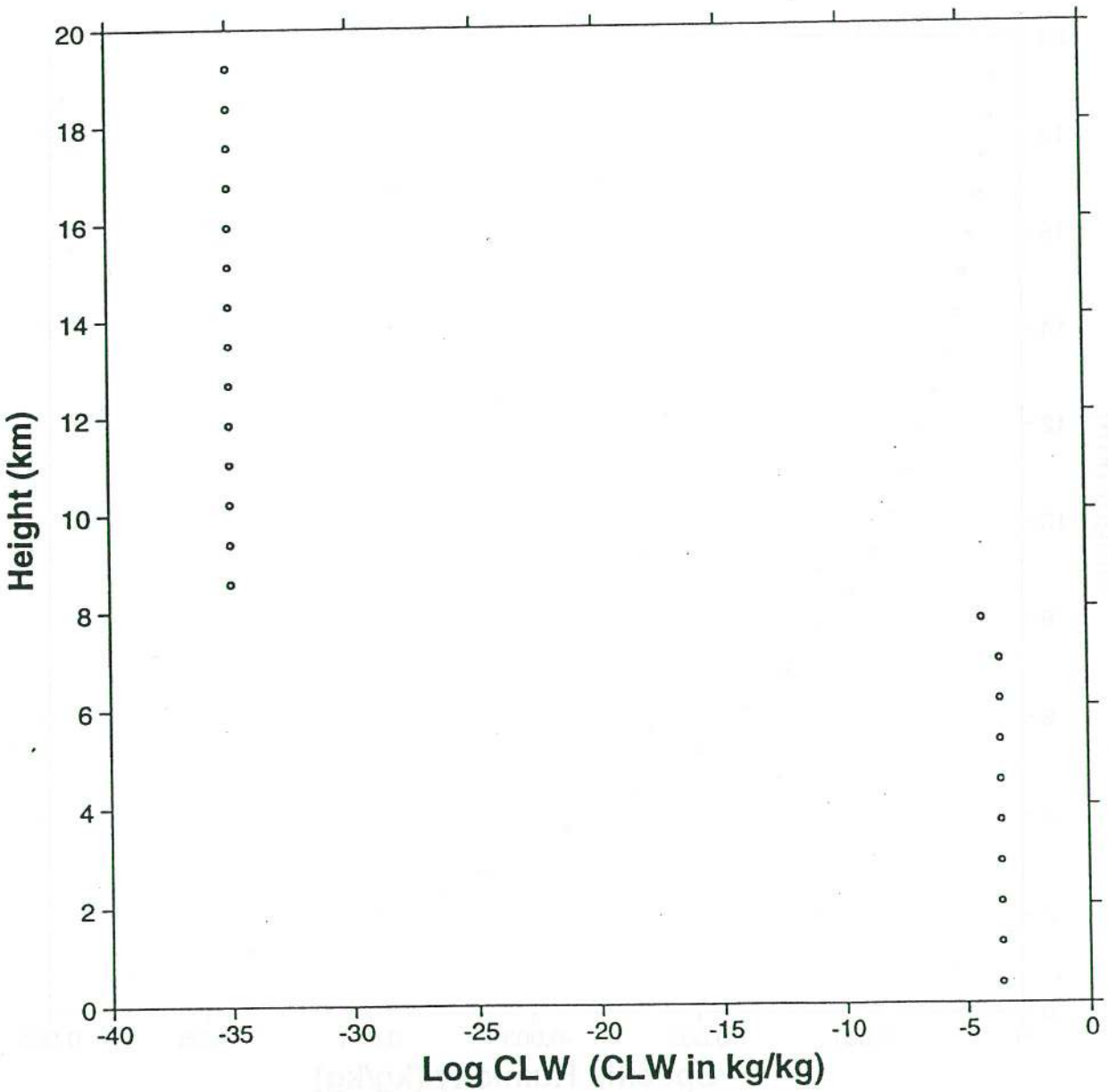
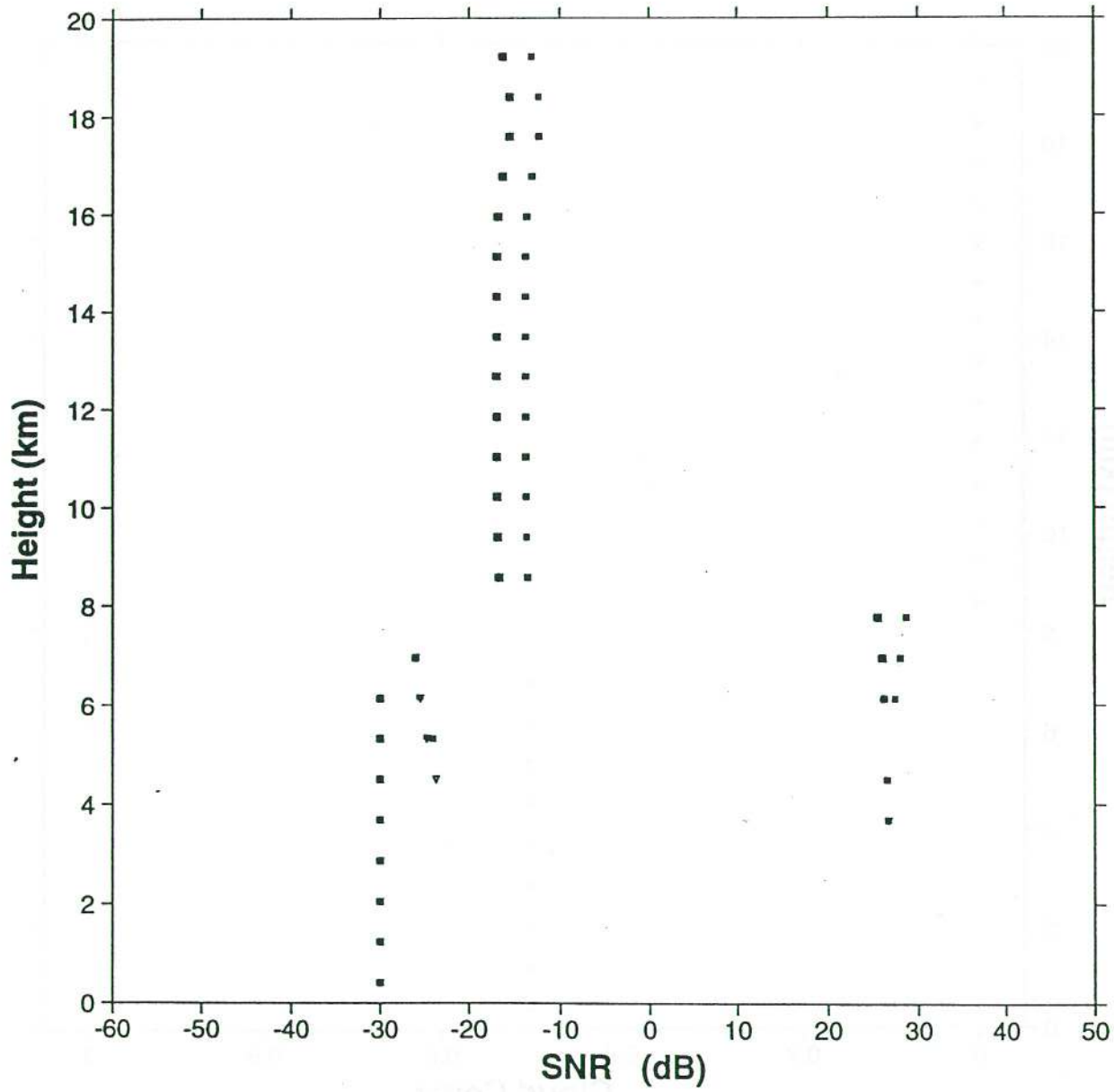


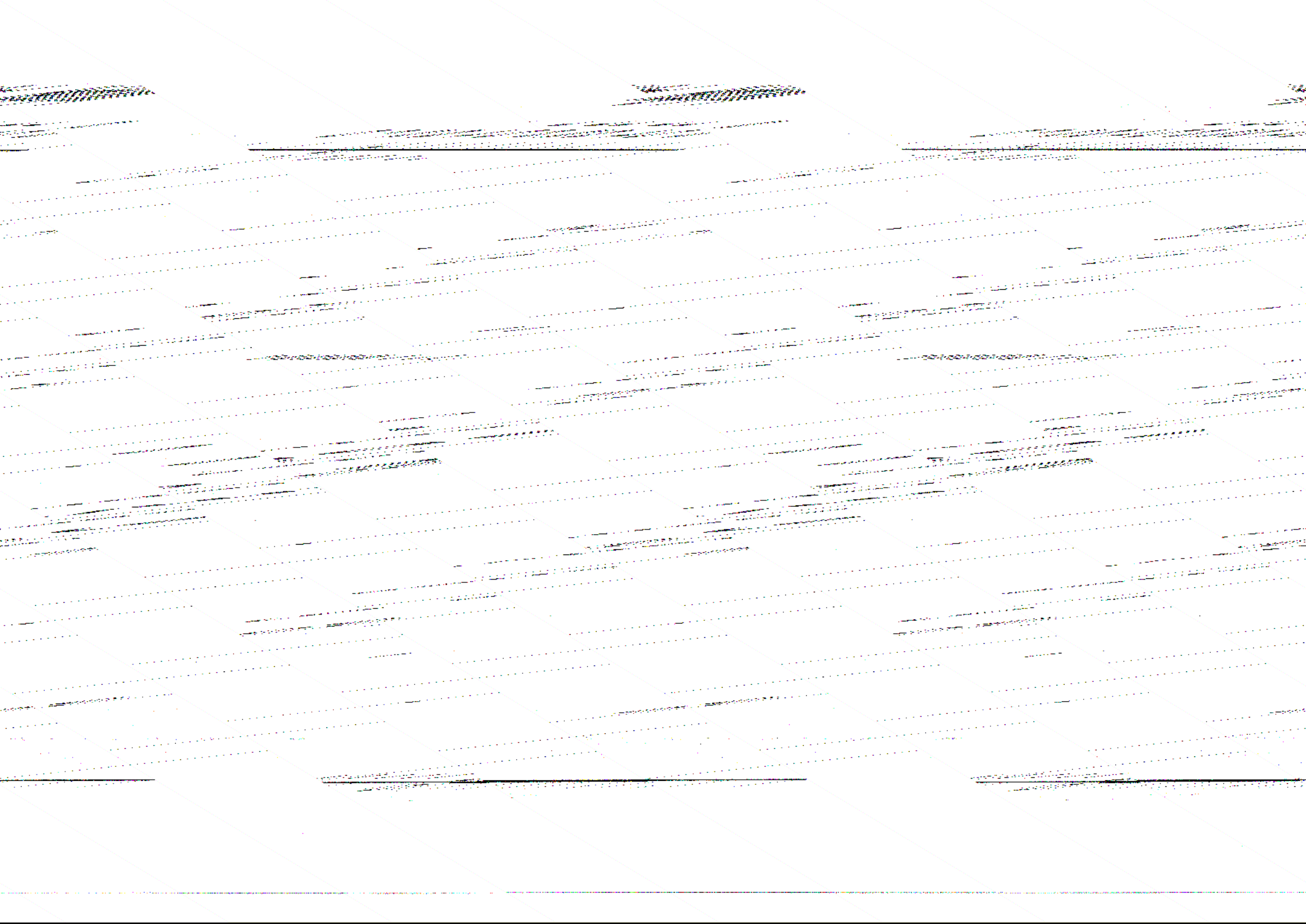
Fig 5.4: Profiles of variables used or computed in the DWL post-processor.
(a) Temperature profile.

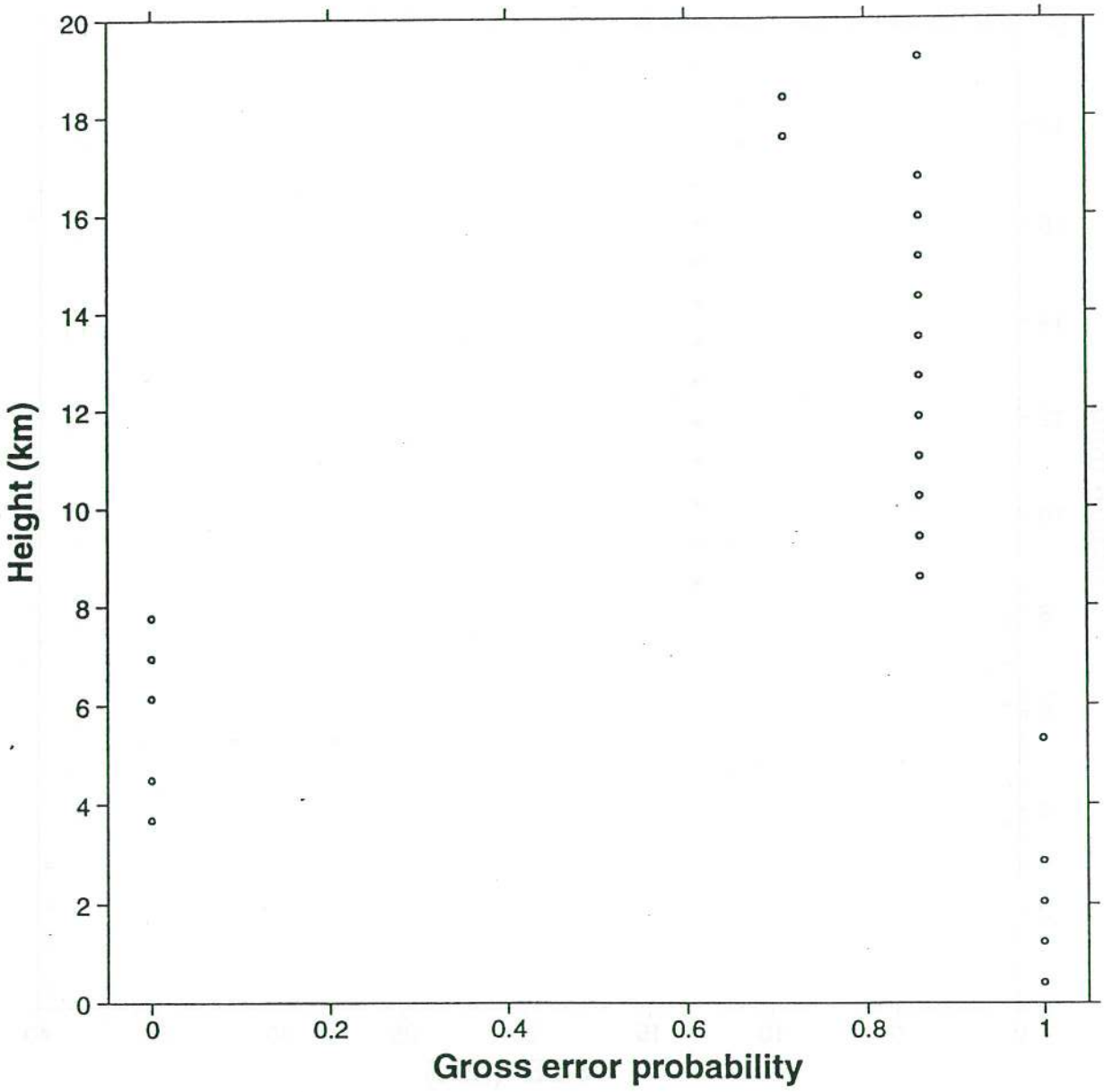


(c) As (a), but for cloud liquid water.

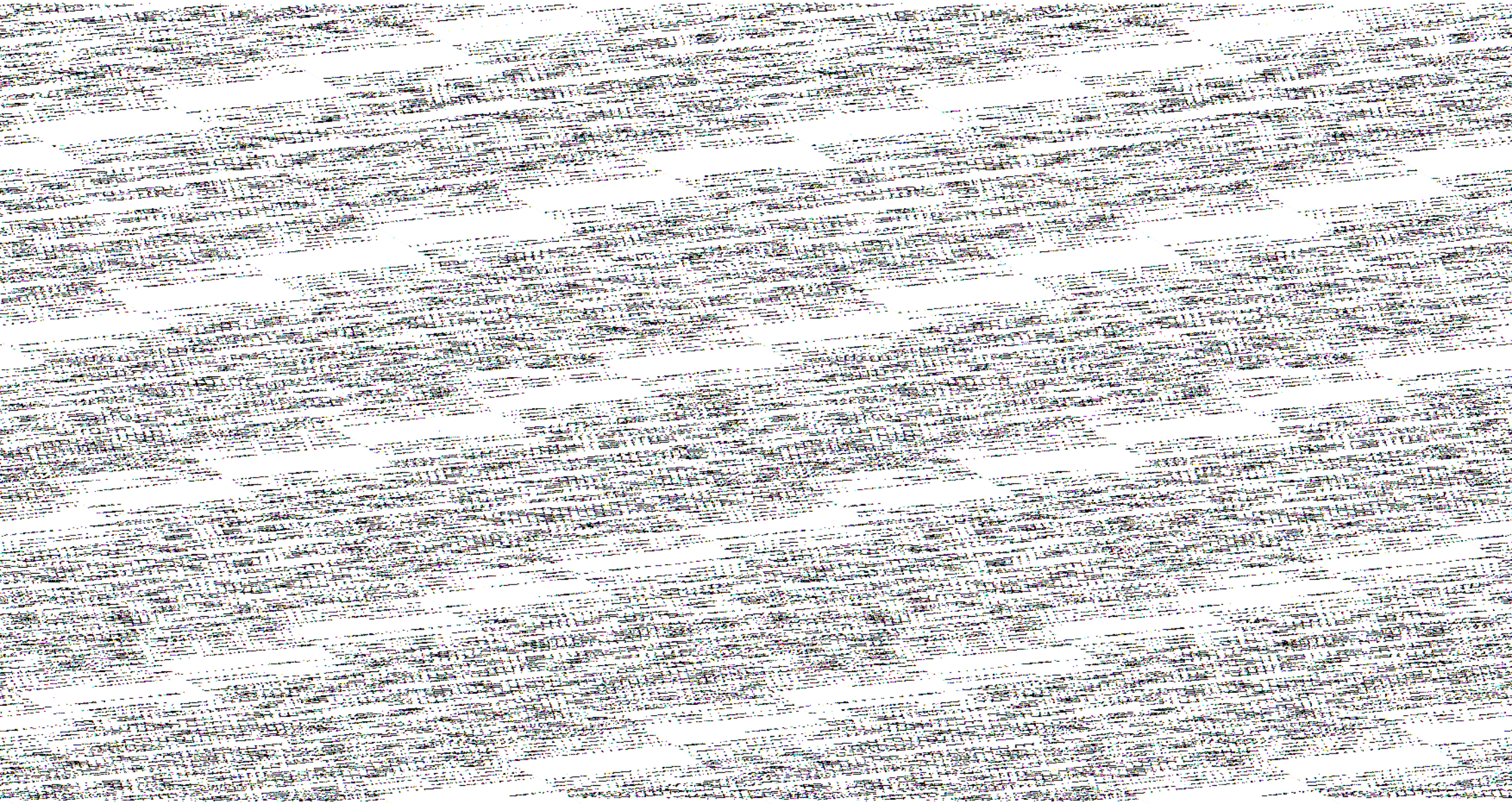


(e) As (a), but for processed SNR at 10 different scans and the integrated value (solid square; right-most symbol). The scenario of orbit II and instrument I was used.





(g) As (a), but for P_G.



The signal-to-noise ratio of DWL measurements depends mainly on the aerosol backscatter coefficient and the transmission from the backscattering volume to the instrument. Reduced signal-to-noise ratio will result in a decreased accuracy of the retrieved wind, and ultimately in a gross error, i.e. a failure to detect the proper Doppler shifted frequency. The number of gross errors will be significant for DWL LOS winds, and it will depend on instrument and orbit specifications. The management of gross errors is of the utmost importance for the use of any data in NWP data assimilation, and therefore an assessment of the quantity of gross errors is used to simulate this type of error for all data types.

The technical steps undertaken to generate the database, the "nature run", the data base and their consistency were validated. The output BUFR files contain the statistical data used to generate the noisy observations, and therefore the subsequent use of the database is made very flexible. Before performing OSSEs, further studies on the contents of the data base should be conducted, specifically to compare the distribution and quality of the DWL measurements in the three simulated scenarios of most interest (1-3). Studies on DWL LOS wind processing algorithms could be carried out with the high resolution data set (scenario 5).

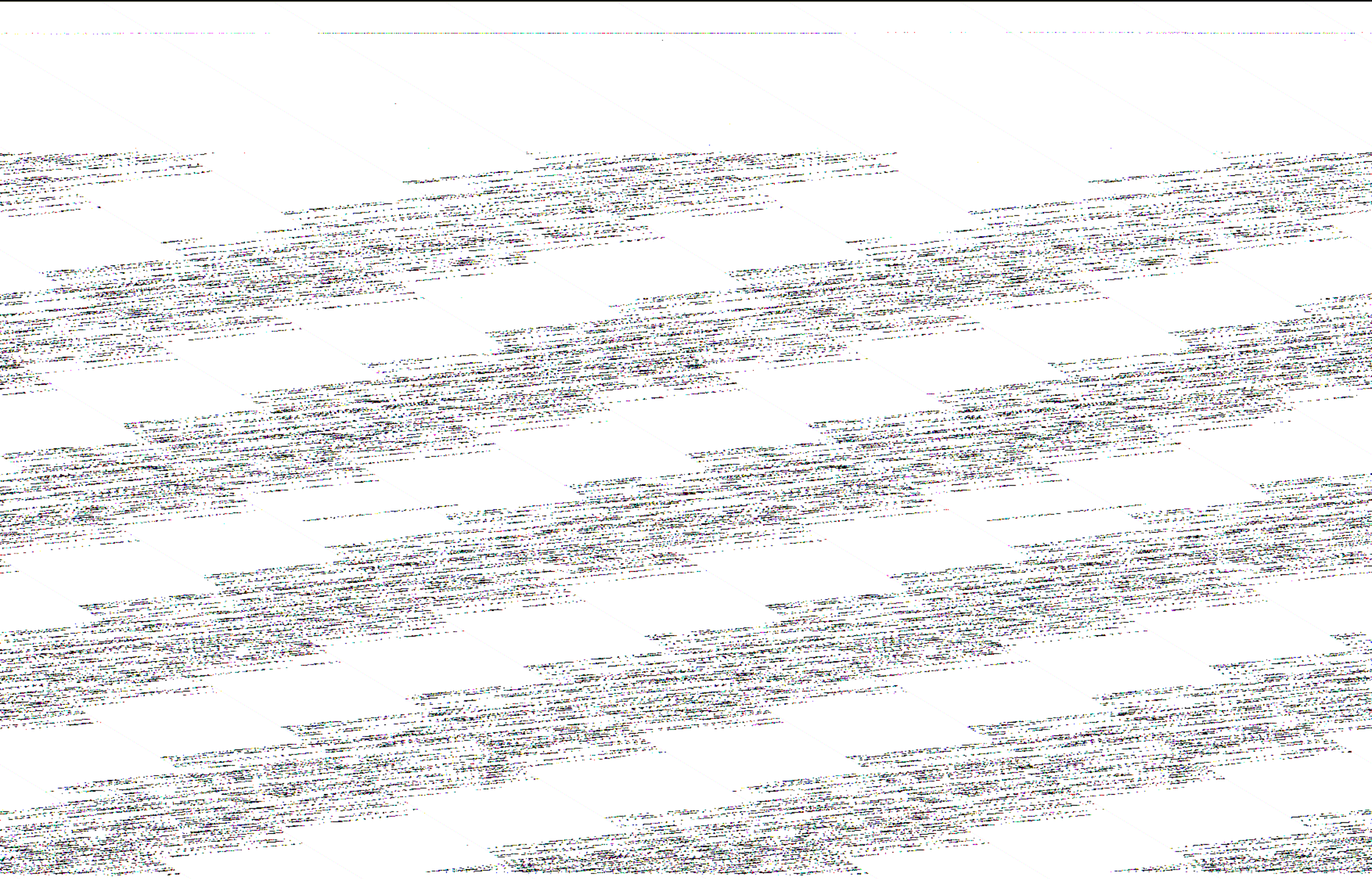
The simulation data base includes distributions of all conventional and real-time available satellite observations with realistic error distributions. The database may be very useful to study the characteristics of data assimilation schemes, and see whether systematic analysis errors develop during assimilation.

Acknowledgements

In the observation simulation experiment the scientific assistance of Pierre Flamant, Alain Dabas and Frédéric Lieutaud from LMD and Andrew Lorenc and Richard Graham from the UK Met. Office has been very useful. The time and efforts from Chris Readings (ESTEC) and Eva Oriol (ESRIN) are also appreciated. At ECMWF numerous people have contributed, especially Philippe Courtier. Anders Person provided us with a professional description of the meteorological conditions in the period of interest (section 3). The resources for the work were provided by ESA and ECMWF.

References

Betout, P, D Burridge and Ch Werner, 1989. Doppler Lidar Working Group Report, ESA SP-1112 (ISBN 92-9092-096-3), ed. T D Guyenne, ESA Publications Division, ESTEC, Noordwijk, The Netherlands.



APPENDIX A: COMPUTATION OF STANDARD DEVIATION AND GROSS ERROR RATE FOR THREE RANGE GATE AVERAGE LOS WIND ERRORS

For the simulation of global DWL data at the coarse resolution, the individual range gates are represented by 1 in 3 points at roughly 1 km vertical resolution. To obtain the general standard deviation of error and gross error probability statistics at this reduced resolution, we simulated the actual processing of three observations at each value of expected SNR. Given the expected SNR value for the central range gate, we simulate a value of measured SNR from an exponential distribution (*Courtier et al.*, 1992) and compute a noisy LOS wind speed V_o using P_G and σ from Figure 4.2. We computed the maximum likelihood estimate for the "true" value, V_P, given three noisy observations V_o^1, V_o^2 and V_o^3 simulated from a "true" value V_T. The following equation gives the probability density for the estimated value V according to Bayes theorem:

$$p(V \mid V_o^1 \cap V_o^2 \cap V_o^3) = p(V_o^1 \mid V)\, p(V_o^2 \mid V)\, p(V_o^3 \mid V)\, p(V) / p(V_o^1 \cap V_o^2 \cap V_o^3),$$

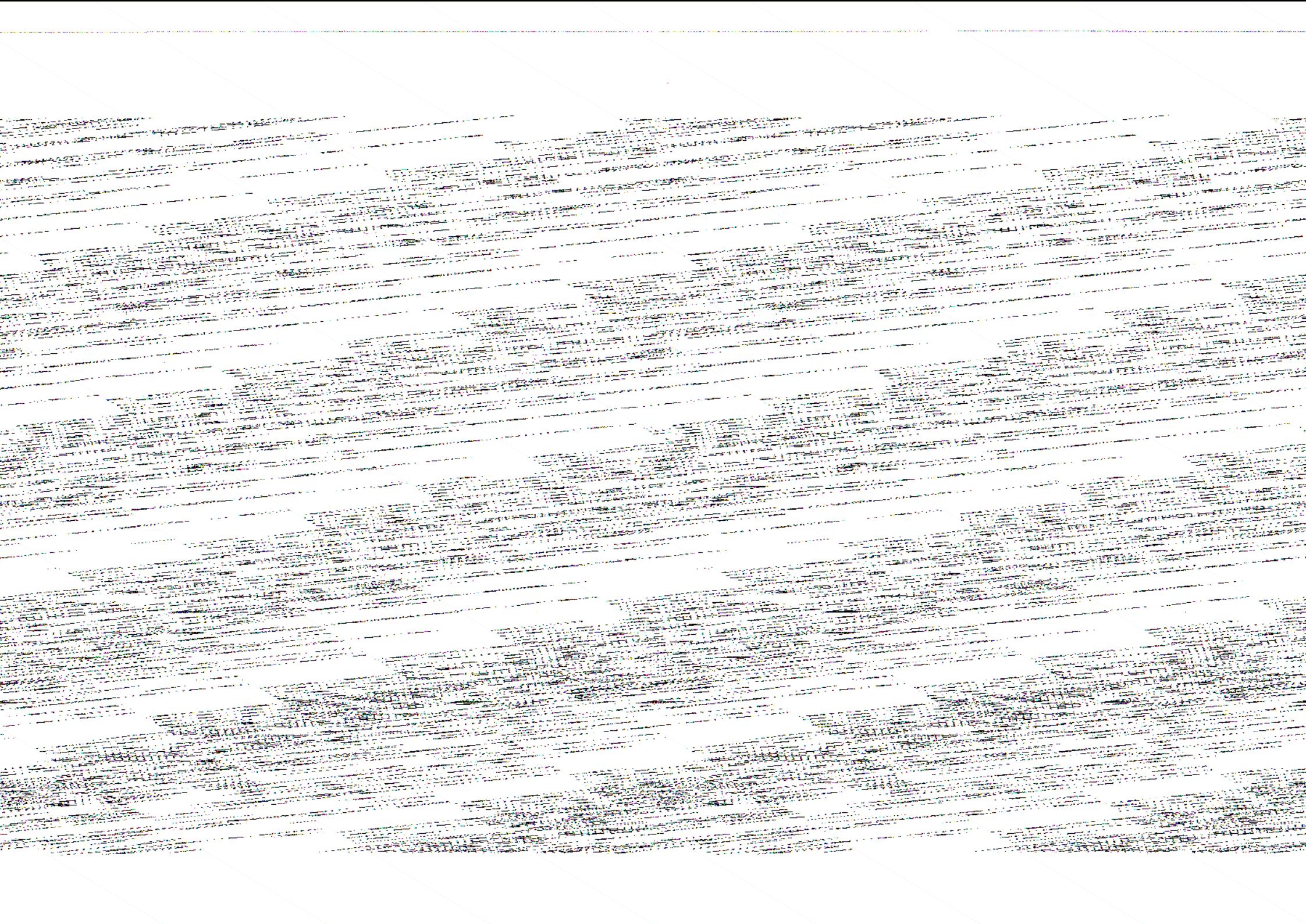
where

$$p(V_o^1 \cap V_o^2 \cap V_o^3) = \int_V p(V_o^1 \mid V)\, p(V_o^2 \mid V)\, p(V_o^3 \mid V)\, p(V)\, dV$$

is the prior probability of the observed set of values V_o^1, V_o^2 and V_o^3, and constant in the exercise of finding V_P. p(V) is the prior probability of an observed value V, and it is assumed to vary insignificantly when varying V. To find V_P we search the entire speed domain. Subsequently, we evaluate the distribution of processed estimates of V_P, given a true value for V, V_T:

$$p(V_P \mid V_T) = \int_V \int_V \int_V p(V_P \mid V_o^1 \cap V_o^2 \cap V_o^3)\, p(V_o^1 \mid V_T)\, p(V_o^2 \mid V_T)\, p(V_o^3 \mid V_T)\, dV_o^1\, dV_o^2\, dV_o^3$$

Here, we replaced $p(V_P \mid V_o^1 \cap V_o^2 \cap V_o^3)$ with the Kronecker δ-function $\delta(V_P)$, since a particular combination of V_o^1, V_o^2 and V_o^3 will result in a unique processed value V_P. In last instance, the acquired distribution $p(V_P \mid V_T)$ was fitted by a probability distribution with a normally distributed and constant part. These fits were reasonable (not shown). The speckle effect was simulated (exponential SNR distribution) using 6 samples of three range gate combinations at each (expected) SNR value. The solid lines in Figure 4.4 represent the reference P_G and σ. The speckle effect increases the gross error rate above -11.5 dB, and reduces P_G and σ for values below -11.5 dB. The results were smoothed with a 5 point filter, and are shown in Figure 4.5.



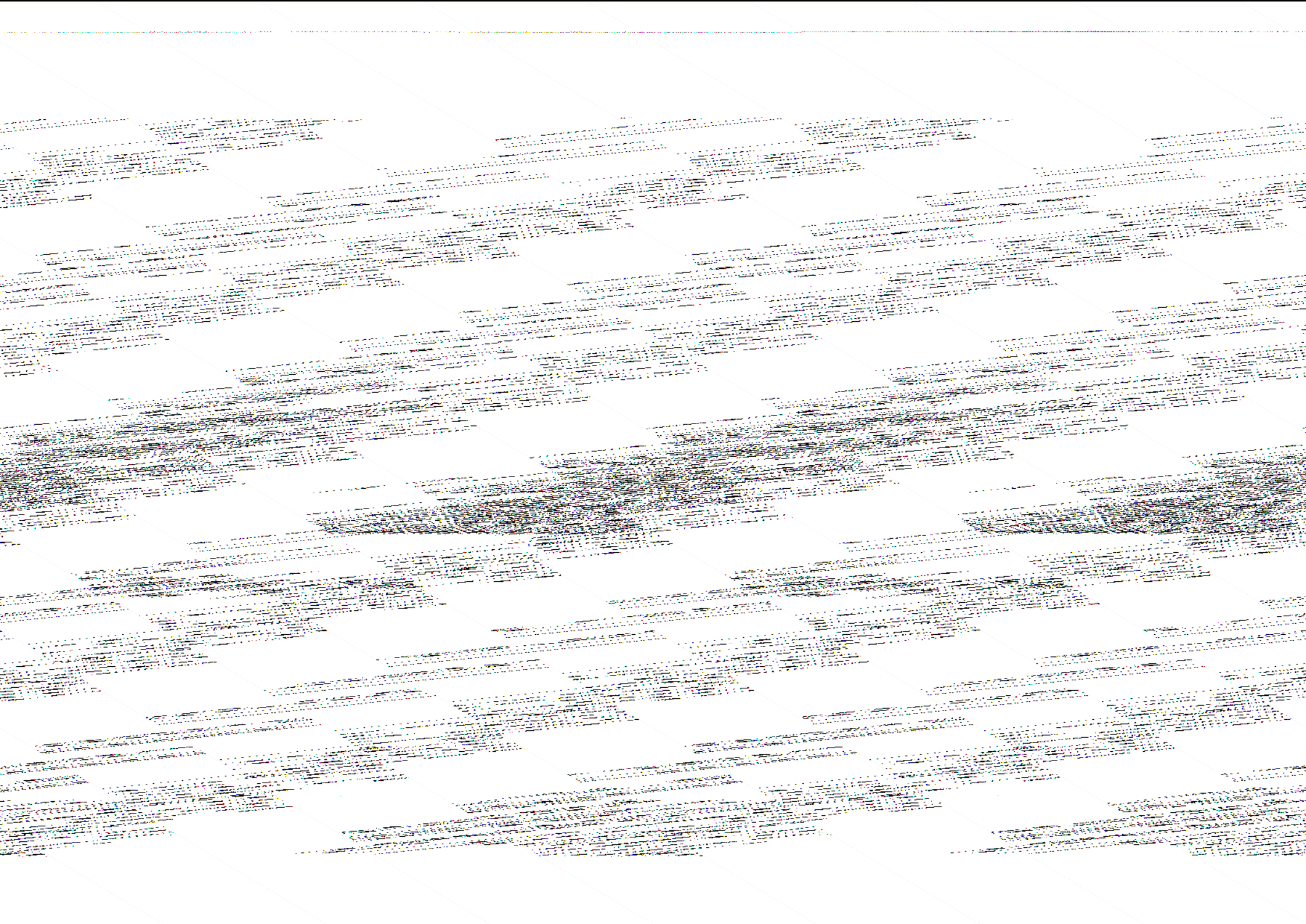
APPENDIX C: BUFR FORMATS

In BUFR, a message is defined as a byte sequence, starting with the four letters "BUFR" and ending with "7777". In the database, each message contains a number of reports of the same type, so that the maximum length of the expanded message is constant (40 000 words). Element descriptors are used to relate a string of bits to the value of a certain parameter (*WMO manual on codes*). Each element descriptor, indicated by a number starting with "0", represents a fixed number of bits in the data section, and provides a unique scaling and reference value to interpret such a set of bits. A list of descriptors therefore uniquely describes the contents of the binary data section. A fixed sequence of descriptors can be replaced by a sequence descriptor, indicated by a number starting with "3". Repetition of descriptors is indicated by a leading descriptor with a number equal to 100000 + X*1000 + Y, where X is the number of following descriptors to be replicated and Y denotes the number of times. If Y = 0 then the next element descriptor would be "(Extended) delayed descriptor replication factor", and have the Y value (in the data section). Operator descriptors start with "2" and can be used for instance to denote that values follow which relate to previously-defined elements. A reference bitmap is then set up to denote to which of the previous descriptors they refer. In our case related values are first order statistics (standard deviation of error σ and range of gross error R), gross error probability P_G and gross error indicator, i.e. the variables that determine the noisy value, given the simulated "true" value. In BUFR, missing data values are denoted by setting all bits to one.

New BUFR element descriptors had to be added to be able to store the data compatible with the BUFR format. These new elements are listed in Table C.1.

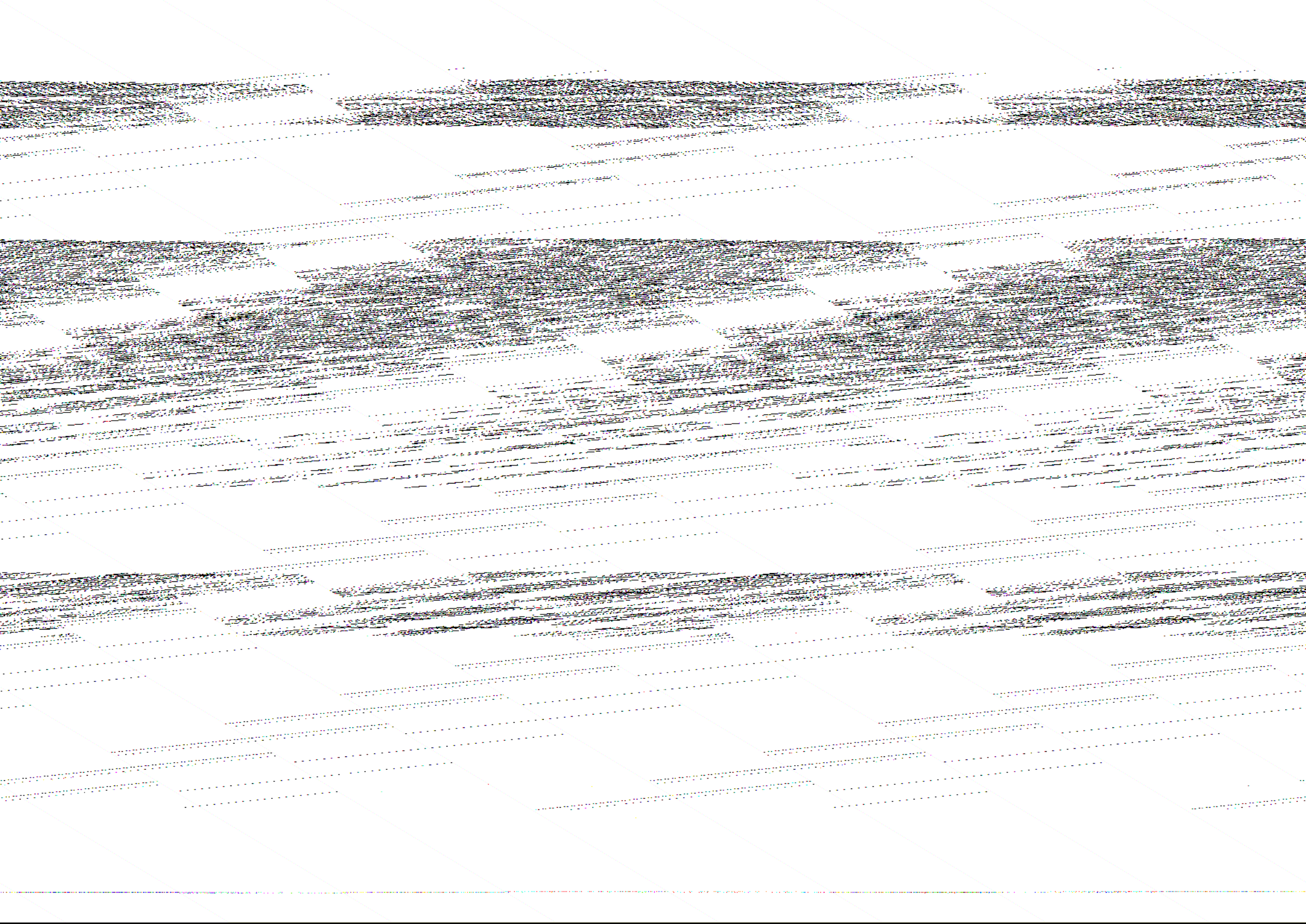
Table C.1: New BUFR element descriptors.

Descriptor number	Descriptor name	Units	Bits	Accuracy	Reference
033241	Gross error probability	-	10	0.001	0.
033242	Gross error indicator	code	2		
001205	Satellite identifier	code	10		
002201	Simulated satellite instrument data used in processing	code	4		
011201	Horizontal line of sight component	m/s	13	0.1	-409.6
013201	Cloud liquid water	kg/kg	14	10^{-7}	0.



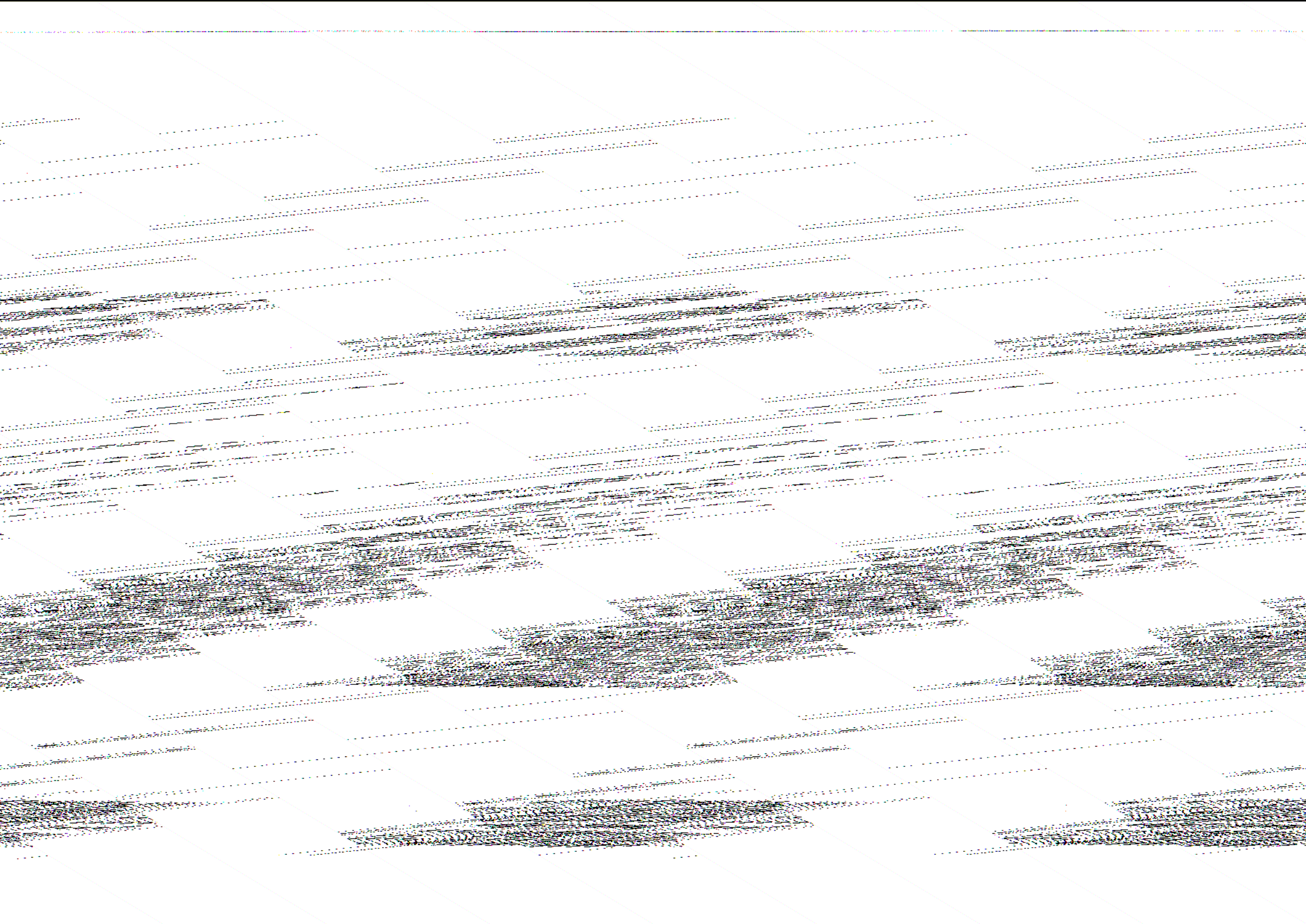
C.1 Land SYNOP

Descriptor number	Descriptor name	Units
001001	wmo block number	numeric
001002	wmo station number	numeric
002001	type of station	code table
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
005001	latitude (high accuracy)	degree
006001	longitude (high accuracy)	degree
007001	height of station	m
010004	pressure	Pa
010051	pressure reduced to mean sea level	Pa
010061	3 hour pressure change	Pa
010063	characteristic of pressure tendency	code table
011011	wind direction at 10 m	degree tru
011012	wind speed at 10 m	m/s
012004	dry bulb temperature at 2m	K
012006	dew point temperature at 2m	K
013003	relative humidity	%
020001	horizontal visibility	m
020003	present weather	code table
020004	past weather (1)	code table
020005	past weather (2)	code table
020010	cloud cover (total)	%
008002	vertical significance (surface observation)	code table
020011	cloud amount	code table
020013	height of base of cloud	m
020012	cloud type	code table
020012	cloud type	code table
020012	cloud type	code table
008002	vertical significance (surface observation)	code table
020011	cloud amount	code table
020012	cloud type	code table
020013	height of base of cloud	m
008002	vertical significance (surface observation)	code table
020011	cloud amount	code table
020012	cloud type	code table
020013	height of base of cloud	m
008002	vertical significance (surface observation)	code table
020011	cloud amount	code table
020012	cloud type	code table
020013	height of base of cloud	m
008002	vertical significance (surface observation)	code table
020011	cloud amount	code table
020012	cloud type	code table
020013	height of base of cloud	m
013023	total precipitation past 24 hours	kg/m**2
013013	total snow depth	m
222000	quality information follow	
031031	data present indicator	numeric
--------	(49 times) ---------	---------



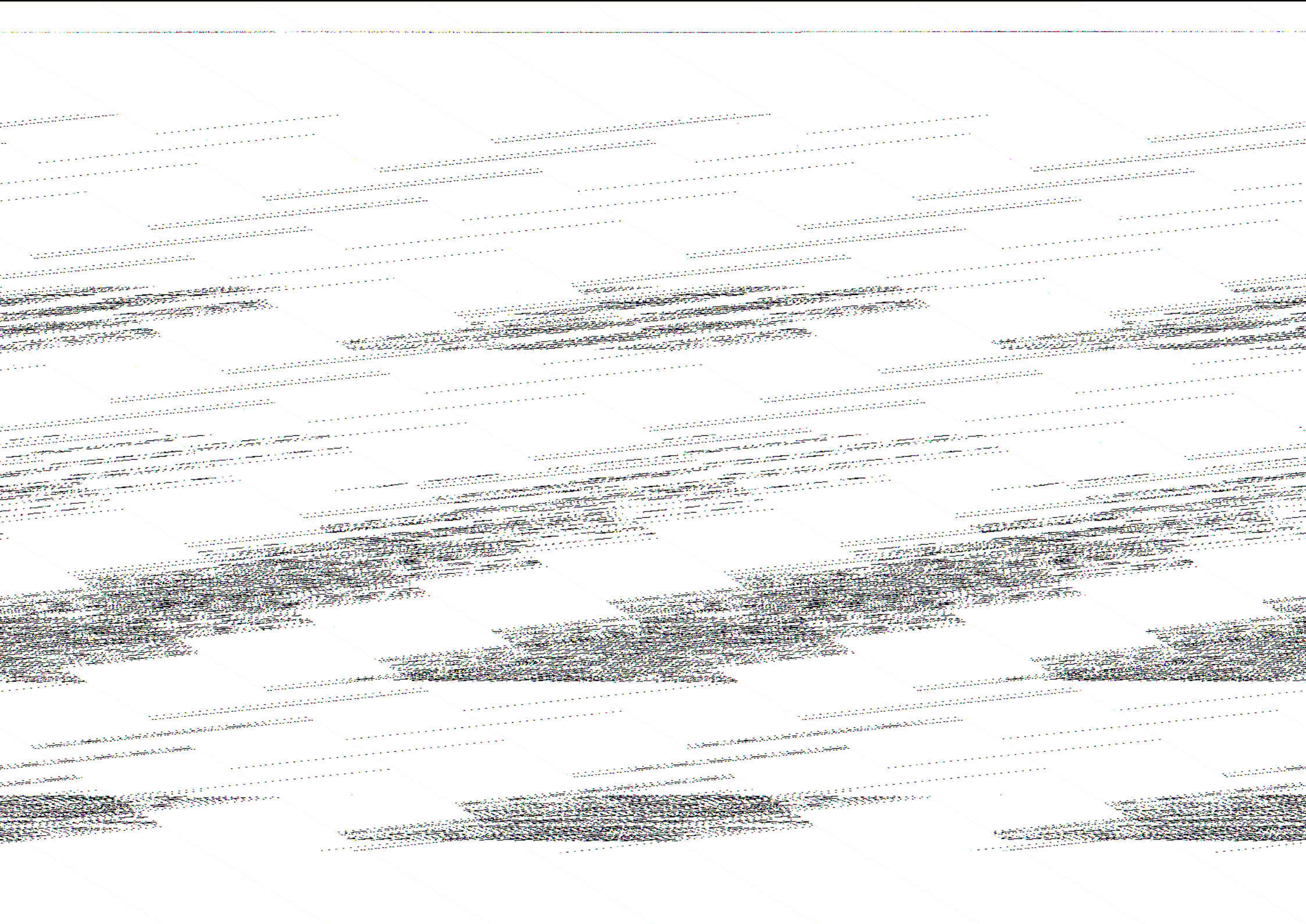
C.2 SHIP

Descriptor number	Descriptor name	Units
001011	ship's call sign	ccittia5
001012	direction of motion of moving observing	degree tru
001013	speed of motion of moving observing plat	m/s
002001	type of station	code table
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
005002	latitude (coarse accuracy)	degree
006002	longitude (coarse accuracy)	degree
010004	pressure	Pa
010051	pressure reduced to mean sea level	Pa
010061	3 hour pressure change	Pa
010063	characteristic of pressure tendency	code table
011011	wind direction at 10 m	degree tru
011012	wind speed at 10 m	m/s
012004	dry bulb temperature at 2m	K
012006	dew point temperature at 2m	K
013003	relative humidity	%
020001	horizontal visibility	m
020003	present weather	code table
020004	past weather (1)	code table
020005	past weather (2)	code table
020010	cloud cover (total)	%
008002	vertical significance (surface observation)	code table
020011	cloud amount	code table
020013	height of base of cloud	m
020012	cloud type	code table
020012	cloud type	code table
020012	cloud type	code table
022042	sea temperature	K
012005	wet bulb temperature at 2m	K
222000	quality information follow	
031031	data present indicator	numeric
--------	(33 times) ---------	---------
001031	generating centre	code table
001201	generating application	code table
033007	% confidence	numeric
--------	(30 times) ---------	---------
235000	cancel backward data reference	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
010004	pressure	Pa
011192	u - component at 10 m	m/s
011193	v - component at 10 m	m/s
012004	dry bulb temperature at 2m	K
013003	relative humidity	%
224000	first order statistics follow	
236000	backward reference bit map	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric



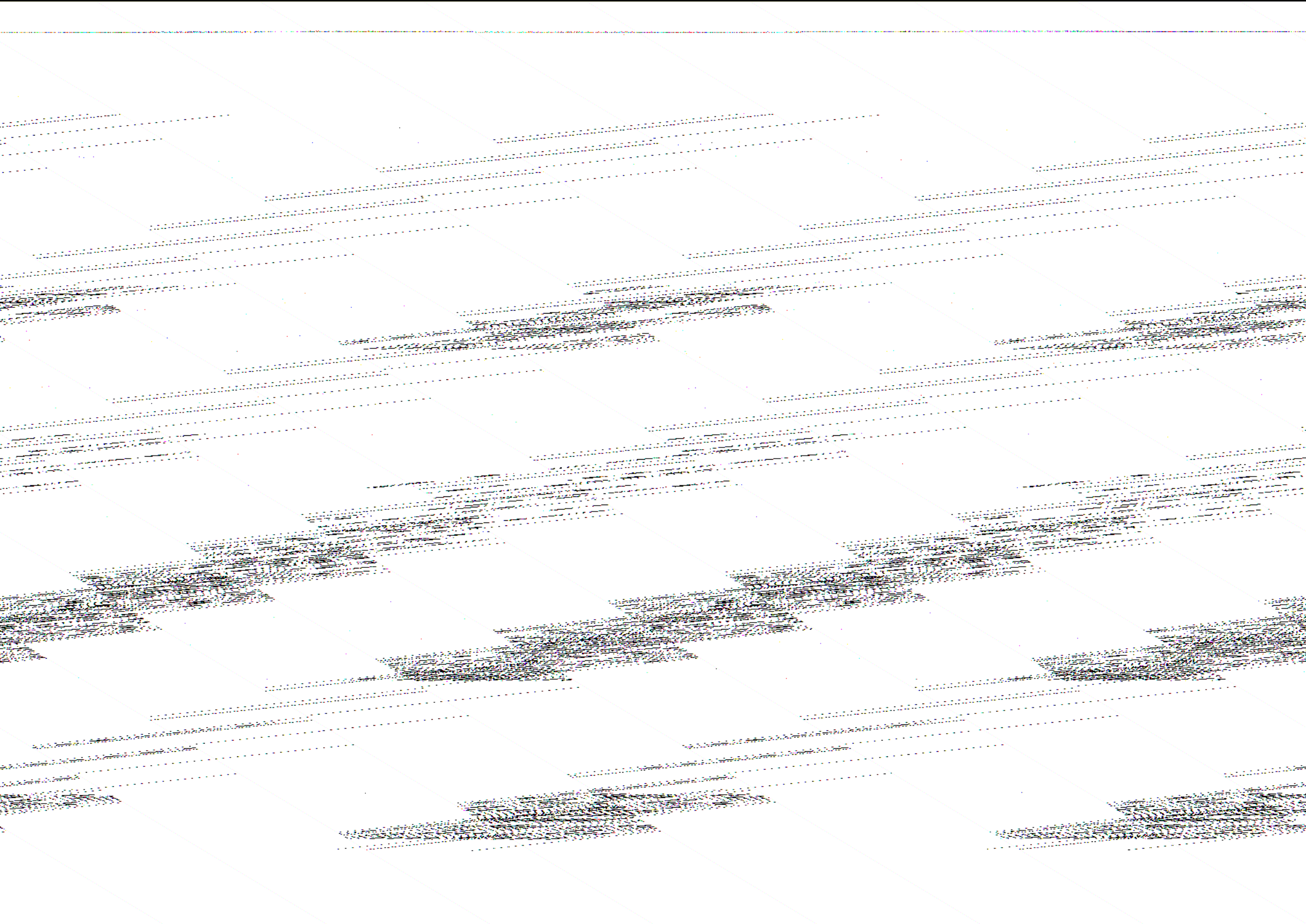
C.3 BUOY

Descriptor number	Descriptor name	Units
001005	buoy/platform identifier	numeric
001012	direction of motion of moving observing	degree tru
001013	speed of motion of moving observing plat	m/s
002001	type of station	code table
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
005002	latitude (coarse accuracy)	degree
006002	longitude (coarse accuracy)	degree
010004	pressure	Pa
010051	pressure reduced to mean sea level	Pa
010061	3 hour pressure change	Pa
010063	characteristic of pressure tendency	code table
011011	wind direction at 10 m	degree tru
011012	wind speed at 10 m	m/s
012004	dry bulb temperature at 2m	K
012006	dew point temperature at 2m	K
013003	relative humidity	%
020001	horizontal visibility	m
020003	present weather	code table
020004	past weather (1)	code table
020005	past weather (2)	code table
020010	cloud cover (total)	%
008002	vertical significance (surface observation)	code table
020011	cloud amount	code table
020013	height of base of cloud	m
020012	cloud type	code table
020012	cloud type	code table
020012	cloud type	code table
022042	sea temperature	K
222000	quality information follow	
031031	data present indicator	numeric
---	(32 times) ---	---
001031	generating centre	code table
001201	generating application	code table
033007	% confidence	numeric
---	(14 times) ---	---
235000	cancel backward data reference	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
010004	pressure	Pa
011192	u - component at 10 m	m/s
011193	v - component at 10 m	m/s
012004	dry bulb temperature at 2m	K
013003	relative humidity	%
224000	first order statistics follow	
236000	backward reference bit map	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric



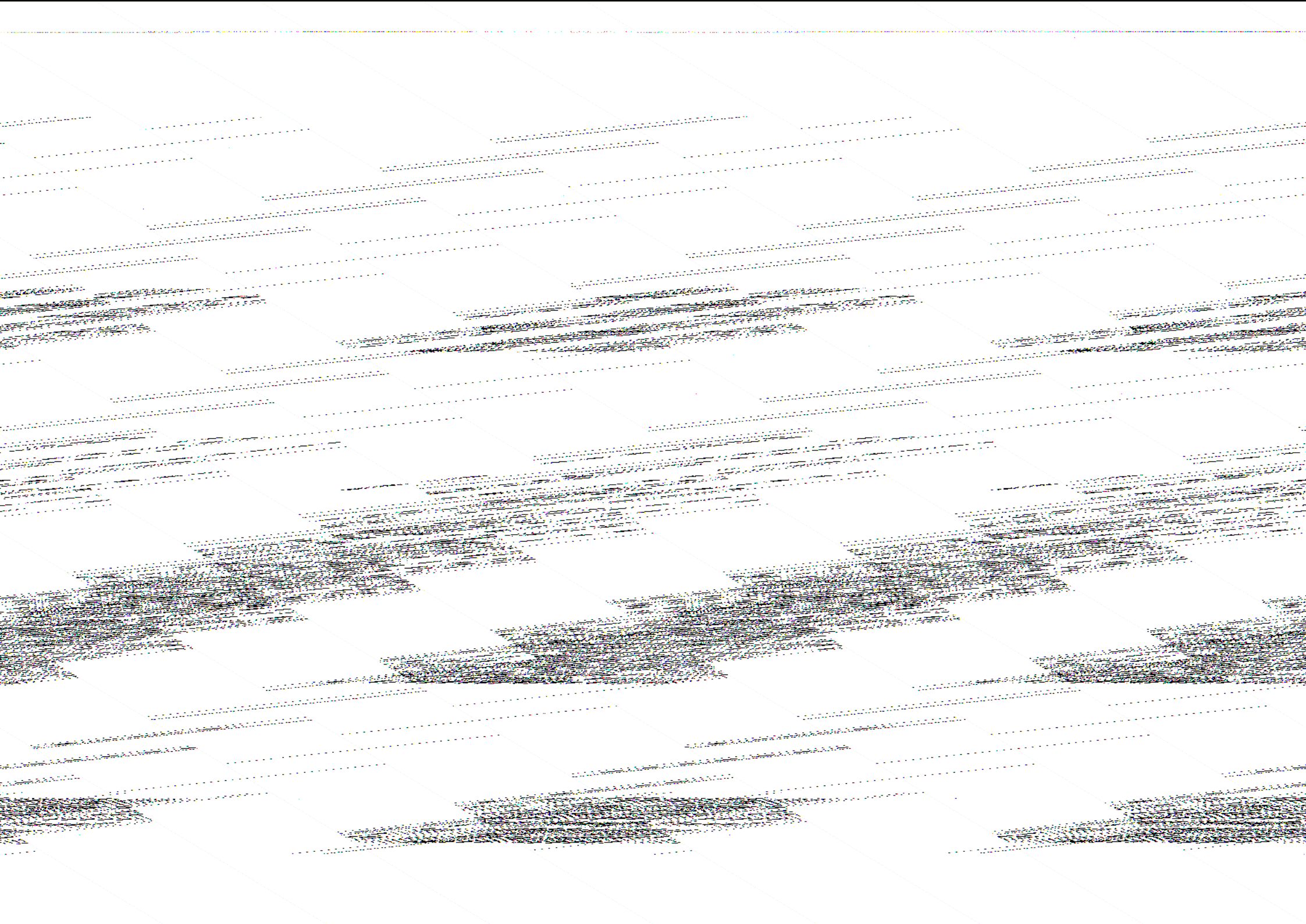
C.4 AIREP

Descriptor number	Descriptor name	Units
001006	aircraft identifier	ccittia5
002061	aircraft navigational system	code table
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
005001	latitude (high accuracy)	degree
006001	longitude (high accuracy)	degree
008004	phase of aircraft flight	code table
007002	height or altitude	m
012001	temperature/dry bulb temperature	K
011001	wind direction	degree tru
011002	wind speed	m/s
011031	degree of turbulence	code table
011032	height of base of turbulence	m
011033	height of top of turbulence	m
020041	airframe icing	code table
222000	quality information follow	
031031	data present indicator	numeric
---	(18 times) ---	---
001031	generating centre	code table
001201	generating application	code table
033007	% confidence	numeric
---	(12 times)	---
235000	cancel backward data reference	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
007004	pressure	Pa
011003	u-component	m/s
011004	v-component	m/s
012001	temperature/dry bulb temperature	K
013003	relative humidity	%
224000	first order statistics follow	
236000	backward reference bit map	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
---	(5 times) ---	---
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric
224255	first order statistics value marker	
---	(5 times) ---	
224000	first order statistics follow	
237000	use previously defined bit map	
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric
224255	first order statistics value marker	
---	(5 times) ---	



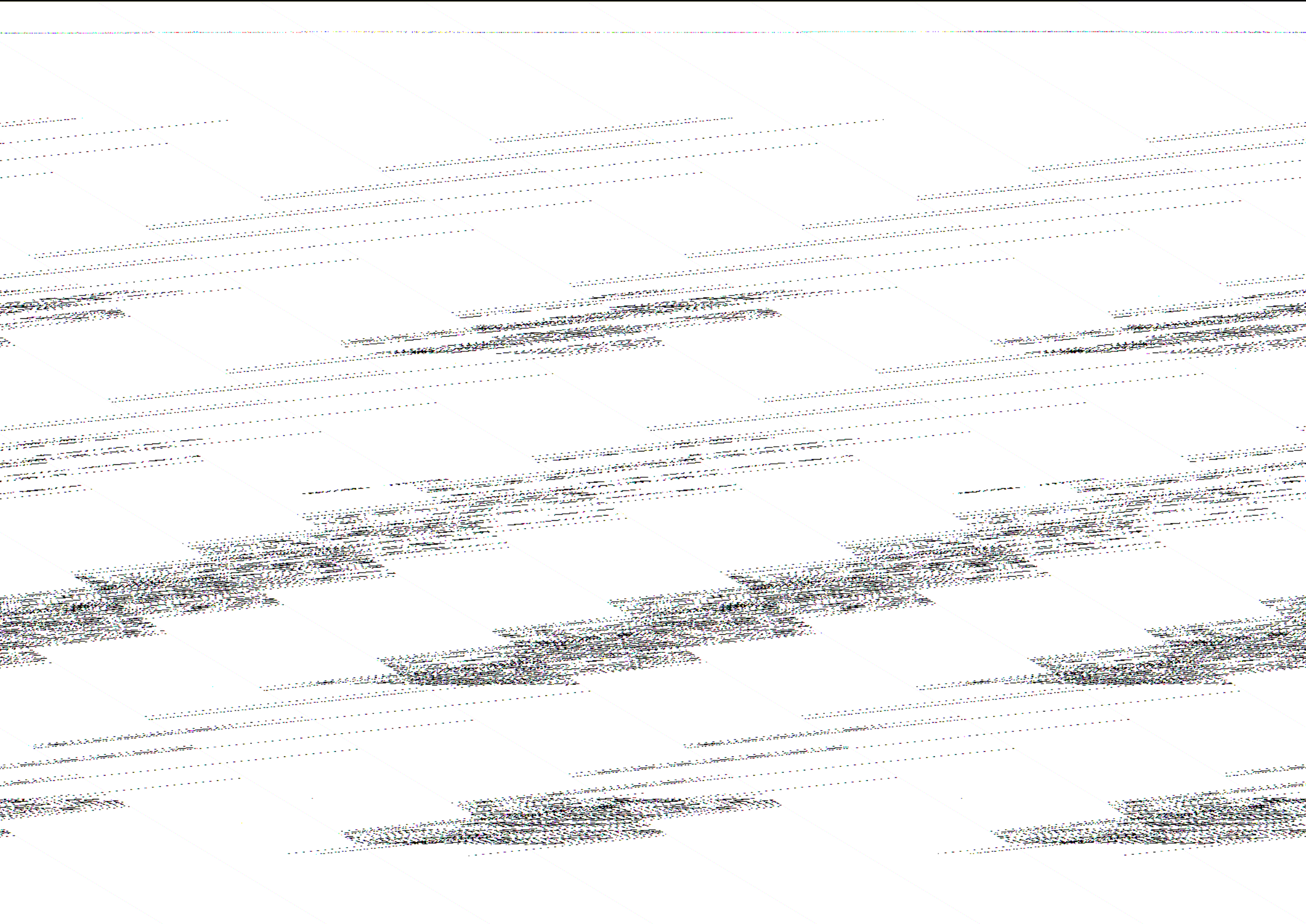
C.5 TEMP

Descriptor number	Descriptor name	Units
001001	wmo block number	numeric
001002	wmo station number	numeric
002011	radiosonde type	code table
002012	radiosonde computational method	code table
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
005001	latitude (high accuracy)	degree
006001	longitude (high accuracy)	degree
007001	height of station	m
020010	cloud cover (total)	%
008002	vertical significance (surface observation)	code table
020011	cloud amount	code table
020013	height of base of cloud	m
020012	cloud type	code table
020012	cloud type	code table
020012	cloud type	code table
031001	delayed descriptor replication factor	numeric
007004	pressure	Pa
008001	vertical sounding significance	flag table
010003	geopotential	m**2/s**2
012001	temperature/dry bulb temperature	K
012003	dew point temperature	K
011001	wind direction	degree tru
011002	wind speed	m/s
----------	(31 times the last 7 parameters)	
222000	quality information follow	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
----------	(238 times) --------	---------
001201	generating application	code table
031002	extended delayed descriptor replication	numeric
033007	% confidence	numeric
----------	(238 times)	---------
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
007004	pressure	Pa
010003	geopotential	m**2/s**2
012001	temperature/dry bulb temperature	K
013003	relative humidity	%
011003	u-component	m/s
011004	v-component	m/s
----------	(31 times the last 6 parameters)	
224000	first order statistics follow	
236000	backward reference bit map	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
----------	(186 times) --------	---------
001031	generating centre	code table
001032	generating application	code table



C.6 PILOT

Descriptor number	Descriptor name	Units
001001	wmo block number	numeric
001002	wmo station number	numeric
002011	radiosonde type	code table
002012	radiosonde computational method	code table
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
005001	latitude (high accuracy)	degree
006001	longitude (high accuracy)	degree
007001	height of station	m
031001	delayed descriptor replication factor	numeric
007004	pressure	Pa
008001	vertical sounding significance	flag table
010003	geopotential	m**2/s**2
011001	wind direction	degree tru
011002	wind speed	m/s
---	(31 times the last 5 parameters)	
222000	quality information follow	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
---	(168 times) --------	---
001031	generating centre	code table
001201	generating application	code table
031002	extended delayed descriptor replication	numeric
033007	% confidence	numeric
---	(168 times)	---
235000	cancel backward data reference	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
007004	pressure	Pa
011003	u-component	m/s
011004	v-component	m/s
---	(31 times the last 3 parameters)	
224000	first order statistics follow	
236000	backward reference bit map	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
---	(93 times) ---------	---
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric
224255	first order statistics value marker	
---	(93 times) ----------	
224000	first order statistics follow	
237000	use previously defined bit map	
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table

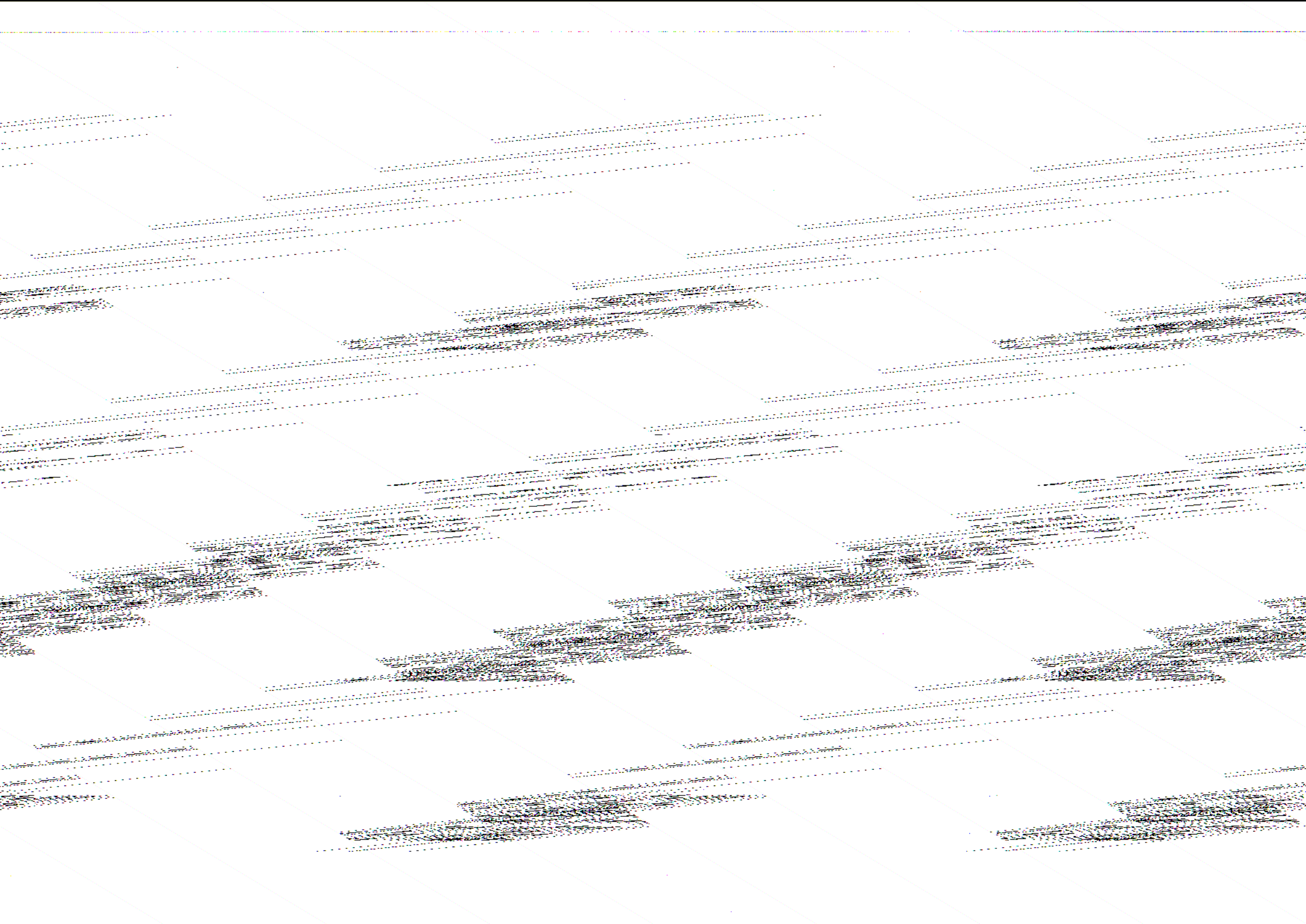


C.7 PAOB

Descriptor number	Descriptor name	Units
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
005001	latitude (high accuracy)	degree
006001	longitude (high accuracy)	degree
010051	pressure reduced to mean sea level	Pa
010004	pressure	Pa
010004	pressure	Pa
012007	virtual temperature	K
222000	quality information follow	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
---------	(11 times) ---------	---------
001031	generating centre	code table
001201	generating application	code table
031002	extended delayed descriptor replication	numeric
033007	% confidence	numeric
---------	(11 times)	---------
235000	cancel backward data reference	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
010004	pressure	Pa
012001	temperature/dry bulb temperature	K
224000	first order statistics follow	
236000	backward reference bit map	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
---------	(twice) ----------	---------
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric
224255	first order statistics value marker	
---------	(twice) ----------	
224000	first order statistics follow	
237000	use previously defined bit map	
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric
224255	first order statistics value marker	
---------	(twice) ----------	
222000	quality information follow	
237000	use previously defined bit map	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
033241	gross error probability	numeric
---------	(twice) ----------	---------

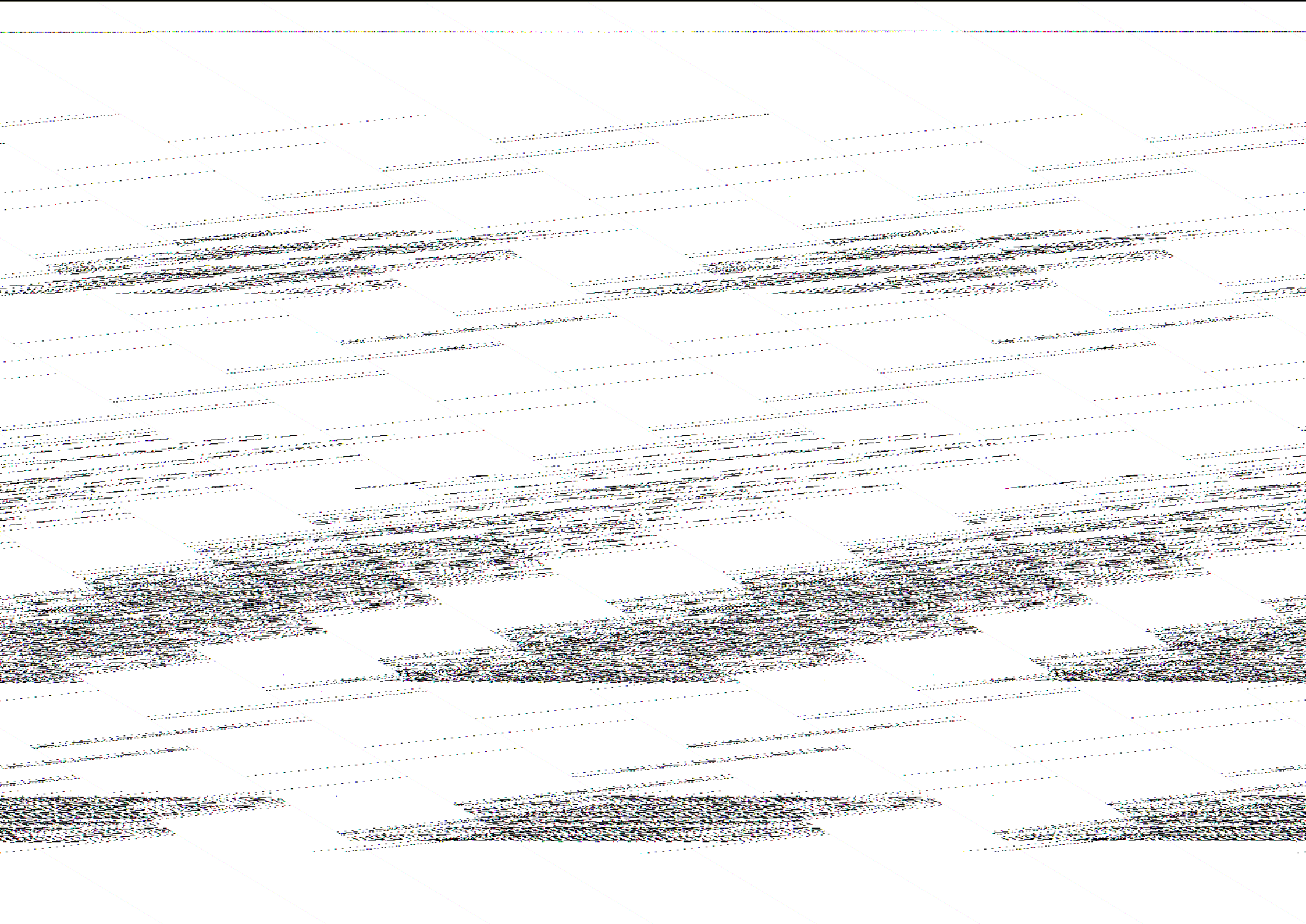
C.8 SATOB

Descriptor number	Descriptor name	Units
001007	satellite identifier	code table
002023	cloud motion computational method	code table
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
004006	second	second
005001	latitude (high accuracy)	degree
006001	longitude (high accuracy)	degree
008003	vertical significance (satellite observation)	code table
010004	pressure	Pa
012001	temperature/dry bulb temperature	K
011001	wind direction	degree tru
011002	wind speed	m/s
222000	quality information follow	
031031	data present indicator	numeric
---	(15 times) ---	---
001031	generating centre	code table
001201	generating application	code table
033007	% confidence	numeric
---	(15 times) ---	---
235000	cancel backward data reference	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
011003	u-component	m/s
011004	v-component	m/s
010004	pressure	Pa
012001	temperature/dry bulb temperature	K
224000	first order statistics follow	
236000	backward reference bit map	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
---	(4 times) ---	---
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric
224255	first order statistics value marker	
---	(4 times) ---	
224000	first order statistics follow	
237000	use previously defined bit map	
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric
224255	first order statistics value marker	
---	(4 times) ---	
222000	quality information follow	
237000	use previously defined bit map	
001031	generating centre	code table

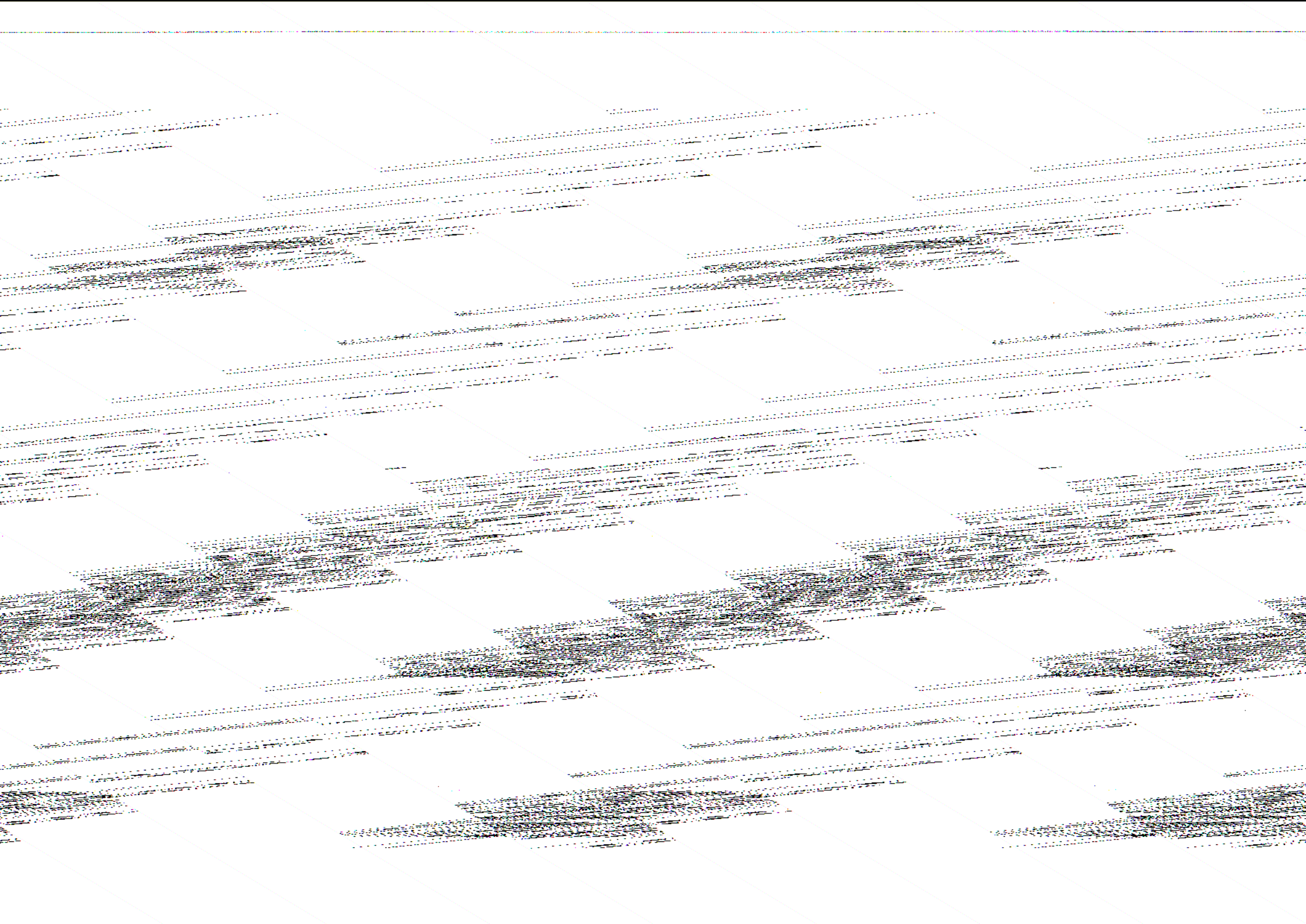


C.9 TOVS

Descriptor number	Descriptor name	Units
001007	satellite identifier	code table
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
004006	second	second
005002	latitude (coarse accuracy)	degree
006002	longitude (coarse accuracy)	degree
007022	solar elevation	degree
002025	satellite channel(s) used in computation flag table	
002022	satellite data processing technique used flag table	
027020	satellite location counter	numeric
008003	vertical significance (satellite observation)	code table
015001	ozon	dobson
008003	vertical significance (satellite observation)	code table
010004	pressure	Pa
020010	cloud cover (total)	%
008003	vertical significance (satellite observation)	code table
008012	land/sea qualifier	code table
010001	height of land surface	m
012061	skin temperature	K
010004	pressure	Pa
008003	vertical significance (satellite observation)	code table
007004	pressure	Pa
007004	pressure	Pa
012007	virtual temperature	K
----------	(15 times the last 3 parameters)	
008003	vertical significance (satellite observation)	code table
007004	pressure	Pa
007004	pressure	Pa
013016	precipitable water	kg/m**2
----------	(3 times the last 3 parameters)	
008003	vertical significance (satellite observation)	code table
012062	equivalent black body temperature	K
----------	(27 times) ----------	-
008003	vertical significance (satellite observation)	code table
010004	pressure	Pa
012001	temperature/dry bulb temperature	K
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
----------	(110 times) --------	-------
001031	generating centre	code table
001201	generating application	code table
031002	extended delayed descriptor replication	numeric
033007	% confidence	numeric
----------	(110 times)	-------
235000	cancel backward data reference	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
012062	equivalent black body temperature	K



Descriptor number	Descriptor name	Units
001205	satellite identifier	code table
002201	simulated satellite instrument	flag table
002022	satellite data processing technique used flag table	
004001	year	year
004002	month	month
004003	day	day
004004	hour	hour
004005	minute	minute
004006	second	second
005002	latitude (coarse accuracy)	degree
006002	longitude (coarse accuracy)	degree
007021	elevation	degree
031002	extended delayed descriptor replication	numeric
007004	pressure	Pa
005021	bearing or azimuth	degree tru
011201	horizontal line of sight component	m/s
---	(Nlev times the last 3 parameters)	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
---	(Nlev x 3 + 12 times)	---
001031	generating centre	code table
001201	generating application	code table
031002	extended delayed descriptor replication	numeric
033007	% confidence	numeric
---	(Nlev x 3 + 12 times)	---
235000	cancel backward data reference	
001031	generating centre	code table
001032	generating application	code table
031002	extended delayed descriptor replication	numeric
007004	pressure	Pa
011201	horizontal line of sight component	m/s
013001	specific humidity	kg/kg
013201	cloud liquid water	kg/kg
020010	cloud cover (total)	%
---	(Nlev times the last 5 parameters)	
224000	first order statistics follow	
236000	backward reference bit map	
031002	extended delayed descriptor replication	numeric
031031	data present indicator	numeric
---	(Nlev x 5 times) ---	---
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric
224255	first order statistics value marker	
---	(Nlev x 5 times) ---	
224000	first order statistics follow	
237000	use previously defined bit map	
001031	generating centre	code table
001032	generating application	code table
008023	first order statistics	code table
031002	extended delayed descriptor replication	numeric



APPENDIX D: ACRONYMS AND ABBREVIATIONS

AIREP	WMO code for single level upper air reports from aircraft or balloon of wind and temperature
AVHRR	Advanced Very High Resolution Radiometer
BUFR	Binary Universal Format Representation
BUOY	WMO code for ocean surface reports of meteorological and oceanographic parameters from a buoy
CAT	Clear Air Turbulence
CMA	Central Memory Array
COF	Comprehensive Observation Format
DWL	Doppler Wind Lidar
ECMWF	European Centre for Medium-range Weather Forecasts
ERS-1	First ESA Remote Sensing Satellite
ESA	European Space Agency
ESRIN	European Space Research INstitute
ESTEC	European Space research and TEchnology Centre
FOV	Field-Of-View
GRIB	GRid In Binary
GTS	Global Telecommunication System
HIRS	High-resolution Infra-red Radiation Sounder
HLOS	Horizontal Line-Of-Sight
IFS	Integrated Forecasting System
INSAT	INdian geostationary SATellite
IR	Infra-Red
LAWS	Laser Atmospheric Wind Sounder
LMD	Laboratoire de Météorologie Dynamique
LOS	Line-Of-Sight
METEOSAT	European geostationary METEOrological SATellite
MSU	Microwave Sounding Unit
NESDIS	National Environmental Satellite, Data and Information Service (of NOAA)
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration (of the USA)
NWP	Numerical Weather Prediction
OI	Optimal Interpolation
OSSE	Observation System Simulation Experiment
PAOB	WMO code for bogus surface pressure observations derived from imagery and ancillary information
PBL	Planetary Boundary Layer (of the atmosphere)
PILOT	WMO code for vertical sounding reports of wind only
PREOB	Observation processor for CMA format
QC	Quality Control
RS	Remote Sensing
RTTOV	Software package for fast radiative transfer computations for TOVS channels
SATOB	WMO code used for satellite data, including cloud-tracked winds
SD	Standard Deviation
SH	Southern Hemisphere
SHIP	WMO code for ocean surface reports of meteorological parameters from a moving platform
SNR	Signal-to-Noise Ratio
SONDE	TEMP or PILOT reports
SSM/I	Special Sensor Microwave Imager
SST	Sea Surface Temperature
SSU	Stratospheric Sounding Unit

APPENDIX E: EXTERNAL ORGANISATION OF DATABASE AND MEANS OF ACCESS

The database is stored for ESA at ECMWF. The format of the database is BUFR for simulated observations and GRIB for the nature run fields produced by the model. To obtain the database, users have to apply to ESA, and after ESA's acceptation ECMWF will provide access. ECMWF can provide users with BUFR and GRIB decoding libraries, plus an example of specific software for decoding the OSSE database and the nature run fields, on the basis of a standard agreement on the provision of ECMWF software.

E.1 **Simulated observations**

The simulated observations are stored in sets of daily binary BUFR files, the files for day D containing all observations between 00:30 UT at day D and 00:30 UT at day D+1. In the ECMWF storage system (ECFILE), the files are kept on the path :

/RDX/DWL_OSSE/

The different observation types are grouped in the set of BUFR files:

- file **bufrsynopYYMMDD** for the date **YYMMDD**, containing land SYNOPs, SHIPs, BUOYs, AIREPs, TEMPs, PILOTs, PAOBs and SATOBs (approximate size : 2.5 Mbyte).

- file **bufrsatemYYMMDD**, containing TOVSs (approximate size : 6.5 Mbyte).

- file **bufIdwlJYYMMDD**, containing DWLs for instrument scenario **I** and orbit scenario **J** (approximate size for low resolution scenarios : 18 Mbyte). **J** is 1 for the 800 km orbit and 2 for the 525 km orbit. **I** is 1 for the 10 Joule instrument, 2 for the 5 Joule descoped instrument, and 3 for the 10 Joule instrument with 9.5 Hz sampling rate (high resolution data run on 930206).

Each file consists of a set of BUFR messages. As BUFR is a table driven format, three BUFR binary tables are needed to decode the files. These tables are the BUFR tables currently used at ECMWF :

- table B : **B000980201** (classification of elements)
- table C : **C000980201** (text and meaning of code and flag tables)
- table D : **D000980201** (list of common sequences)

A full definition of the BUFR format can be found in the WMO publication No. 306 (Manual on Codes Volume I, Part B).

Table E.1: ECMWF local Code Table 2, Version Number 128 for FM 92-VIII Ext. GRIB

CODE FIGURE	MARS ABBREV.	FIELD	UNITS
129	Z	Geopotential	m^2s^{-2}
130	T	Temperature	K
133	Q	Specific humidity	kg/kg
135	W	Vertical velocity	Pa s^{-1}
138	VO	Vorticity (relative)	s^{-1}
139	ST	Surface temperature	K
140	SSW	Surface soil wetness	m (of water)
141	SD	Snow depth	m (of water)
142	LSP	Large scale precipitation*	*m
143	CP	Convective precipitation*	*m
144	SF	Snow fall*	*m (of water)
145	BLD	Boundary layer dissipation*	*$Wm^{-2}s$
146	SSHF	Surface sensible heat flux*	*$Wm^{-2}s$
147	SLHF	Surface latent heat flux*	*$Wm^{-2}s$
151	MSL	Mean sea level pressure	Pa
152	LNSP	Log surface pressure	-
155	D	Divergence	s^{-1}
157	R	Relative humidity	%
164	TCC	Total cloud cover	(0 - 1)
165	10U	10 metre u	m s^{-1}
166	10V	10 metre v	m s^{-1}
167	2T	2 metre temperature	K
168	2D	2 metre dewpoint temperature	K
170	DST	Deep soil temperature	K
171	DSW	Deep soil wetness	m (of water)
173	SR	Surface roughness	m
174	AL	Albedo	-
176	SSR	Surface solar radiation*	*$Wm^{-2}s$
177	STR	Surface thermal radiation*	*$Wm^{-2}s$
178	TSR	Top solar radiation*	*$Wm^{-2}s$
179	TTR	Top thermal radiation*	*$Wm^{-2}s$
180	EWSS	U-stress*	*$Nm^{-2}s$
181	NSSS	V-stress*	*$Nm^{-2}s$
182	E	Evaporation*	*m(of water)
183	CDST	Climatological deep soil temp.	K
184	CDSW	Climatological deep soil wetness	m (of water)
185	CCC	Convective cloud cover	(0 - 1)
186	LCC	Low cloud cover	(0 - 1)
187	MCC	Medium cloud cover	(0 - 1)
188	HCC	High cloud cover	(0 - 1)
195	LGWS	Lat. comp. of gravity wave stress*	*$Nm^{-2}s$
196	MGWS	Mer. comp. of gravity wave stress*	*$Nm^{-2}s$
197	GWD	Gravity wave dissipation*	*$Wm^{-2}s$
198	SRC	Skin reservoir content	m (of water)
201	MX2T	Maximum temp. at 2m since previous post-processing	K
202	MN2T	Minimum temp. at 2m since previous post-processing	K
205	RO	Runoff*	*m
241	CLWC	Cloud liquid water content	kg/kg
242	CC	Cloud cover	(0 - 1)

* denotes field accumulated since start of forecast.

iv) *Fields on pressure levels in spherical harmonics*:

- code figure: 129 (17 levels)
  130 (17 levels)
  133 (17 levels)
  135 (17 levels)
  138 (17 levels)
  157 (17 levels)
  155 (17 levels)