Real Time Data Quality Control and Monitoring in the New Zealand Meteorological Service *

M. J. UDDSTROM AND D. K. PURNELL

New Zealand Meteorological Service
PO Box 722, Wellington, New Zealand

1 March 1989

ABSTRACT

This paper outlines data error detection, adjustment and monitoring algorithms in use at the New Zealand Meteorological Service.

Operational data compatibility adjustment schemes for radiosonde and drifting buoy measurements are described, and the need for a similar adjustment algorithm for satellite radiance measurements is discussed. A prediction error covariance model, suitable for use in a data assimilation scheme operating on a data sparse region is described, and representative "observation error" statistics provided for data verified against first guess and analysis fields.

1. Introduction

Data quality problems can be classified into two broad categories, errors arising in the measurement process, and those caused during transmission of the measured information. However, observing instruments, although designed to make measurements of some particular variable, will almost always return data that are contaminated with unexpected information (as opposed to noise). This additional, unexpected component in the measurements will introduce an "error" which might also be time dependent. The real time system for detection and correction of "data errors" at the New Zealand Meteorological Service (NZMS) has been designed so as to take account of these general properties of all data.

At the NZMS, two different streams of data are available for use in analysis and forecasting. One stream arrives in coded form via communications lines such as the GTS (or AFTN) or PACNET, while the other, usually data derived by remote sensing methods, is available in physical units from local processing. The type of error detection logic employed in the data quality checking is therefore a function of the data source, but can be characterised by the following chain.

*Presented at the WMO Workshop on Data Quality Control Procedures, Reading, 6–10 March 1989.

- Syntax and consistency checking of coded messages.
- Adjustment of data for known time dependent and independent errors.
- Optimal Interpolation analysis filtering and data quality monitoring.

The question of data quality control and monitoring will be discussed under these three headings.

2. Syntax and Bounds Checking

For data arriving at the NZMS in coded form, the decoding software immediately extracts all information that satisfies the syntax rules for the particular codes relevant to the messages. Accordingly, parts of reports may be decoded while others are not. Messages containing syntax errors are flagged, and, if possible, manually corrected. However, if the error is not obvious, and the observing station is under New Zealand control, a re-transmission of the original message may be requested. The only exceptions to this rule are for upper air data, where, if an error in the syntax is detected, the entire report is ignored (or a re-transmission requested). Parameter bounds checking is performed during the decoding operation.

At regular periods after the synoptic observing times, and before the data are made available to the analysis software, two additional quality control checks are applied. The purpose of the first is to detect errors in station identification numbers. If the decoded data do not match the expected observing network schedule then it is assumed that an error has occurred, the data are rejected, and both the unexpected observations, and those expected, but not received, are logged for further analysis. The second applies only to upper temps data and is a hydrostatic consistency and super-adiabatic layers check. Reports containing such errors may be edited manually and corrections made, if possible, but no automated procedure (such as that described by Collins and Gandin, 1988) is used to resolve "obvious" errors.

3. Data Adjustments

Data adjustment schemes have already been implemented for two types of measurements; radiosonde upper air temperature and height data, and locally derived drifting buoy measurements. An adjustment scheme has also been developed for satellite radiance measurements, but is not yet operational.

a. *Radiosonde data*

Almost all radiosonde data used in the NZMS data assimilation scheme are either of type Philips RS4 or Vaisala RS80 (both navaid and non-navaid models). The Vaisala RS80 data are automatically corrected at the observing station for both short and long wave effects, prior to transmission. However, no such corrections are made to the Philips data.

Consequently, stations flying Philips RS4 sondes produce measurements which contain time dependent biases relative to those produced by stations flying RS80 sondes.

ECMWF radiosonde monitoring statistics (Söderman, pers. comm.), Fig. 1, for a station (93012) flying the Philips RS4 radiosonde indicate that there are two types of error present in the uncorrected thermodynamic data. The first is observing-time dependent, in that the 0000 and 1200 UTC observations have different vertical biases, relative to the interpolated first guess and analysis values. The second relates to the sign of these biases; the 0000 UTC (day time) data show little bias with decreasing pressure, contrary to the expected situation for solar contaminated data, while the 1200 UTC (night time) data have an unexpected negative bias with decreasing pressure.

The majority of the time dependent bias errors can be removed from the RS4 data through application of short wave radiation corrections of the type given in Table 1 (from Uddstrom 1984). Using these corrections, day time observations are reduced to effective night time values, although no account is taken of infrared emission and / or absorption sources, which might contribute additional significant errors (McMillin et al., 1987).

Solar Elevation Angle °	**Pressure Level** 700	500	300	200	100	50	30	20 hPa
15	0.0	0.3	0.1	0.4	0.9	1.3	1.1	
	6	7	17	18	32	58	64	
25	0.2	0.4	0.3	0.5	1.1	1.5	1.7	1.3
	6	9	13	17	35	63	78	92
35	0.2	0.2	0.4	0.8	1.1	1.5	1.7	2.2
	4	7	12	19	41	66	94	107
45	0.2	0.4	0.5	1.1	1.4	1.9	2.3	2.9
	5	7	15	22	43	79	112	136
55	0.0	0.2	0.3	0.7	1.3	1.7	1.7	2.6
	1	4	10	18	40	73	101	129
65	0.0	0.3	0.4	0.8	1.2	1.9	2.0	2.7
	2	3	8	16	40	72	100	139
75	-0.3	0.3	0.7	1.1	1.3	1.6	2.2	2.3
	0	2	8	20	45	74	100	121

TABLE 1: Short wave radiation corrections for temperature and geopotential height, and Philips RS4 stations in the New Zealand region (from Uddstrom, 1984).

However, these corrections alone do not adjust the data for the apparently significant time independent errors noted in Fig. 1. Uddstrom (1989) reports on a dual flight experiment which compared RS4 and RS80 measurements. It was found that there is a significant time independent bias in the data for these instruments (see Table 2) and that the magnitude of this error is very close to that indicated by the ECMWF monitoring statistics. Accordingly, once the RS4 data have been corrected to effective night time values, they are further adjusted by the values given in Table 2 (from Uddstrom 1989), to make them consistent with data originating from Vaisala equipment.

Pressure hPa	Geopotential m	Temperature °C	Relative Humidity %
20	-107	-1.2	
30	-106	-0.3	
50	-93	-2.0	
70	-77	-1.7	
100	-54	-1.5	
150	-37	-1.3	
200	-28	-1.0	
250	-20	-1.1	
300	-14	-1.2	
400	-5	-0.6	4
500	-1	-0.4	4
700	2	-0.4	6
850	3	-0.2	-1
1000	2	0.3	1

TABLE 2: RS4 minus RS80 differences for solar corrected data.

It should be noted that the corrections indicated in Table 2 do not agree with those proposed by Nash and Schmidlin (1987). The reasons for this discrepancy are discussed in Uddstrom (1989).

b. *Drifting buoy data*

A significant source of surface information (pressure, air and sea temperature, and wind speed and direction at one metre) from remote ocean areas in the New Zealand region is provided by drifting buoys. The pressure measurements from these platforms are known to be susceptible to time dependent bias type errors, but because of the paucity of conventional surface data in the region it often proves difficult to establish the quality of buoy measurements. Locally processed drifting buoy data are continually monitored by operational forecasters, and the data compared, where possible, with passing ships or other buoys, and manual MSL pressure analyses. If the data from a particular buoy can be shown to contain a bias component, then the calibration parameters for the buoy in question are altered so as to remove the bias during local processing of the Data Collection and location System (DCS) data stream.

The wind speed and direction measurements from drifting buoys are not currently used in either the manual or machine surface analyses, because their relationship to the 10 m wind is not understood. An experiment will soon be carried out in order to establish an appropriate reduction model (Laing, pers comm.). The results from this experiment, which will compare the 10 m marine wind with buoy reports, will be incorporated into the local processing of drifting buoy data, so that the reported winds more nearly reflect 10 m values.

c. *TOVS radiance observations*

The NZMS currently employs a regression algorithm to retrieve thickness profiles from TOVS data. Accordingly there is no requirement to correct radiances for bias errors. However, it is now clear that, depending on the "ground truth" standard chosen, satellite radiance measurements suffer from air-mass dependent bias errors, relative to current radiative transfer equation simulations (Uddstrom, 1988b; Kelly, 1988; McMillin et al., 1989).

It is anticipated that a physical retrieval algorithm (Uddstrom 1988a) will soon be implemented. Because the information in the radiances is provided to a physical retrieval algorithm through the difference between the satellite measurements and radiances synthesised from some first guess atmosphere it is essential that the errors in the estimated radiances be small, relative to the measurement errors. Uddstrom (1988b) and Kelly (1988) indicate that conventional (the so called *universal* γ, δ) radiance correction techniques will result in serious degradation of the information contained in the satellite measurements.

A radiance correction algorithm has been developed which enables radiances to be corrected as a function of an air mass classification, using techniques developed in Uddstrom (1988b) and McMillin et al. (1989). The need for such a scheme is evident from Fig 2., where, for a selection of NOAA10 TOVS channels, the difference between collocated radiative transfer equation radiances and measured values, the so called δ values, are plotted as a function of an air-mass index (see Uddstrom 1988b). Clearly, in most cases a universal δ value would result in spurious errors in retrievals calculated with a physical estimator.

4. Analysis Data Monitoring

In data rich regions, data quality monitoring can be provided through examination of statistics of differences between the observations and short range forecast first guess fields (valid at the observation time), and the observations and the analysed fields (Hollingsworth et al. 1986). In data sparse areas, and in regions where significant orographic features are not well resolved by the forecast model, the interpretation of the differences between measurements and first guess and analysed values is rather more difficult. If it is assumed that the analysis interpolation operator errors are relatively small, then such differences, which are sometimes defined to be "data errors" are, in reality, the sum of forecast model prediction and observation errors.

a. *Analysis error covariance model*

The NZMS multivariate optimal interpolation analysis scheme utilises an error covariance model of the "first guess" fields, which is particularly satisfactory in data sparse regions, and where the (real) orography may change rapidly, relative to the forecast model's horizontal resolution. This error model is currently specified as the sum of five statistically independent components. The first two are stream-functions, having differing horizontal correlation distances, allowing a variation of the geostrophic horizontal-covariance model with pressure. The third controls orographic lee effects below 700 hPa and the fourth is a velocity-potential component. The inclusion of the fifth term, an ageostrophic geopotential, results in the scheme functioning well at all latitudes, including the tropics. The covariance of each component is then modelled as a product of an isotropic horizontal (Gaussian) covariance and a vertical correlation. Therefore while the horizontal-covariance model of each of the five error components is not allowed any variation with altitude, each of the component covariances may have a different horizontal structure, so a vertically weighted sum of the components does provide a variation of the horizontal-covariance structure with altitude. Further, it is a simple matter to arrange for the first guess error covariance to be significantly larger in regions down stream from substantial orographic features.

b. *Forecast model*

The NZMS primitive equation forecast model follows that of McGregor et al. (1978), except that an advective formulation of the equations (rather than the original flux form) is employed, since high orography is encountered on the model grid (see Fig. 3). The basic model variables are:- the surface pressure, and at each of ten σ levels, the u and v components of the wind, the temperature, and mixing ratio. A staggered Arakawa C grid is used for the horizontal fields. The model is initialised using the Bourke and McGregor (1983) vertical mode scheme, and it runs on a polar stereographic grid, where the distance between grid points is 190.5 km at 60°S and approximately 100 km at the equator. At the lateral boundaries all variables are held constant and the divergence is set to zero.

A physics package, as developed for the Australian Region Primitive Equation model, is included and provides:

- vertical mixing - dependent on Richardson number
- surface stress and friction - proportional to low level stability
- surface energy balance - with a diurnal temperature cycle
- a radiation scheme - moisture dependent
- a convection scheme - based on that of Kuo (1965)

c. *Some monitoring statistics*

To demonstrate the data density over the NZMS regional grid, Fig. 3 includes analyses for 1000 – 500 hPa thickness at 1200 UTC, and MSL pressure at 0000 UTC. Over the entire grid (of 63 by 49 points) at 1200 UTC, there are only seven radiosonde reports. The significance of the local satellite derived measurements is evident. Unfortunately, NOAA derived SATEMs are not currently used. The surface network is much more extensive, as indicated in the 0000 UTC analysis, yet large data voids are present, and the importance of drifting buoy and ship reports obvious. In the data sparse regions, the analysis algorithm will be forced to rely heavily on the forecast model first guess. If the forecast model predictions suffer serious errors, then these will eventually advect into some area where data are encountered and will become apparent as large differences between the observations and the first guess field. In these cases, the "data errors" almost certainly arise from the forecast first guess rather than from the observations.

Notwithstanding these precautionary comments, verification statistics (deduced from an operational verification data base) for an area centred on New Zealand (see Fig. 4), and various data types are displayed in Fig. 5. Both observation minus first-guess (a six hour prognosis) and observation minus analysis statistics are displayed. When compared with the height and wind results given in Hollingsworth et al. (1986), for North America (a homogeneous network), it would appear that the equivalent NZMS observation minus analysis statistics are quite similar. However, the observation minus first guess rms difference statistics for the NZMS verification area are a little poorer than those given by Hollingsworth, a not unexpected result. The substantial errors in the AIREPS are merely an aberration, since they are small sample estimates.

The satellite thickness errors indicate that an important component in the difference statistics arises from a bias in the data. The current regression retrieval scheme assumes the pressure of the lowest radiating surface is uniform over a wide latitude zone (30°N - 30°S, 30° - 60°S, and 60° - 90°S). Therefore the satellite thicknesses will, when the surface pressure is different from the assumed climatology, introduce a perhaps significant bias from this source. Regardless, the utility of the satellite observations is apparent.

The MSL pressure statistics are much as expected. The "bogus" observations show poorest verification errors (against the analyses). However, since these data are digitised from manual analyses, are from data sparse regions, and are given the lowest weight in the analysis algorithm, this result is satisfactory. The cause for the differences between ship and buoy observations is not obvious.

5. Concluding Remarks

The NZMS operates an NWP data assimilation and forecast cycle in a data sparse region. For this reason, observations which are known to contain measuring instrument introduced biases, are adjusted to some standard. Currently, radiosonde and drifting buoy data are "corrected". For satellite radiance data to be used in a physical retrieval algorithm, it is necessary to calibrate the satellite measurements with some radiative transfer equation model and standard "ground truth". To this end, satellite radiance measurements will soon be adjusted according to air-mass type.

In terms of data quality monitoring, it is evident that even in a data sparse region, data may be compared with forecast first guess and analysis fields, and information about data quality and the performance of the data assimilation algorithm inferred. In particular, the observation minus first guess difference statistics are useful in diagnosing forecast model errors and observing network deficiencies, as well as comparing observation errors. However, interpretation of these types of statistics, with respect to the magnitude of observation errors for a particular type of data, is not straight forward.

Acknowledgements. We wish to thank Jim Renwick for providing the verification statistics, and Mark Sinclair for the NWP fields. Also we extend thanks to Mike Revell for contributing information about the prediction model characteristics. Other colleagues have helped in the review process.

References

Bourke, W.P. and J.L. McGregor, 1983: A non-linear vertical mode initialisation scheme for a limited area prediction model. *Mon. Wea. Rev.*, **111**, 1749–1771.

Collins, W.G. and L.S. Gandin, 1988: The hydrostatic checking of radiosonde heights and temperatures. *NOAA/NMC Office Note 344*, 27 pp.

Hollingsworth, A., D.B. Shaw, P. Lönnberg, L. Illari, K. Arpe and A.J. Simmons, 1986: Monitoring of observation and analysis quality by a data assimilation system. *Mon. Wea. Rev.*, **114**, 862–879.

Kelly, G.A., and J-F. Flobert, 1988: Radiance Tuning. In the Technical Proceedings of the *Fourth International TOVS Study Conference*, Pub. CIMSS, University of Wisconsin, 99–117.

Kuo, H.L., 1965: On the formation and intensification of tropical cyclones through latent heat release by cumulus convection. *J. Atmos. Sci.*, **22**, 40–63

McGregor, J.L., L.M. Leslie and D.J Gaunlett 1978: The ANMRC limited-area model: consolidated formulation and operational results. *Mon. Wea. Rev.*, **106**, 427–438.

McMillin, L.M., M.E. Gelman, A. Sanyal and M. Sylva, 1987: A method to use satellite retrievals as a transfer standard to determine systematic radiosonde errors, *Mon. Wea. Rev.*, **116**, 1091–1102.

McMillin, L.M., L.J. Crone and D. Crosby, 1989: Adjusting satellite radiances by regression with an orthogonal transformation to a prior estimate. *J. Appl. Meteor*, in press.

Nash, J. and F.J. Schmidlin, 1987: Final Report of the WMO International Radiosonde Comparison (U.K. 1984, U.S.A. 1985), *Instruments and Observing Methods Rep No. 30*, 103 pp.

Uddstrom, M.J., 1984: Solar radiation corrections for New Zealand and Australian radiosondes, *New Zealand Met. Service, Scientific Rep 6*, 34 pp.

Uddstrom M.J., 1988a: Retrieval of atmospheric profiles from satellite radiance data by typical shape function maximum a posteriori simultaneous retrieval estimators. *J. Appl. Meteor.* **27**, 516–549.

Uddstrom M.J., 1988b: The effect of collocation radiosonde errors on the assessment of the performance of a physical retrieval estimator. In the Technical Proceedings of the *Fourth International TOVS Study Conference*, Pub. CIMSS, University of Wisconsin, 371–389.

Uddstrom M.J., 1989: A comparison of Philips RS4 and Vasisala RS80 Radiosonde Data. *J. Atmos. Oceanic Tech*, **6**, 201–207.

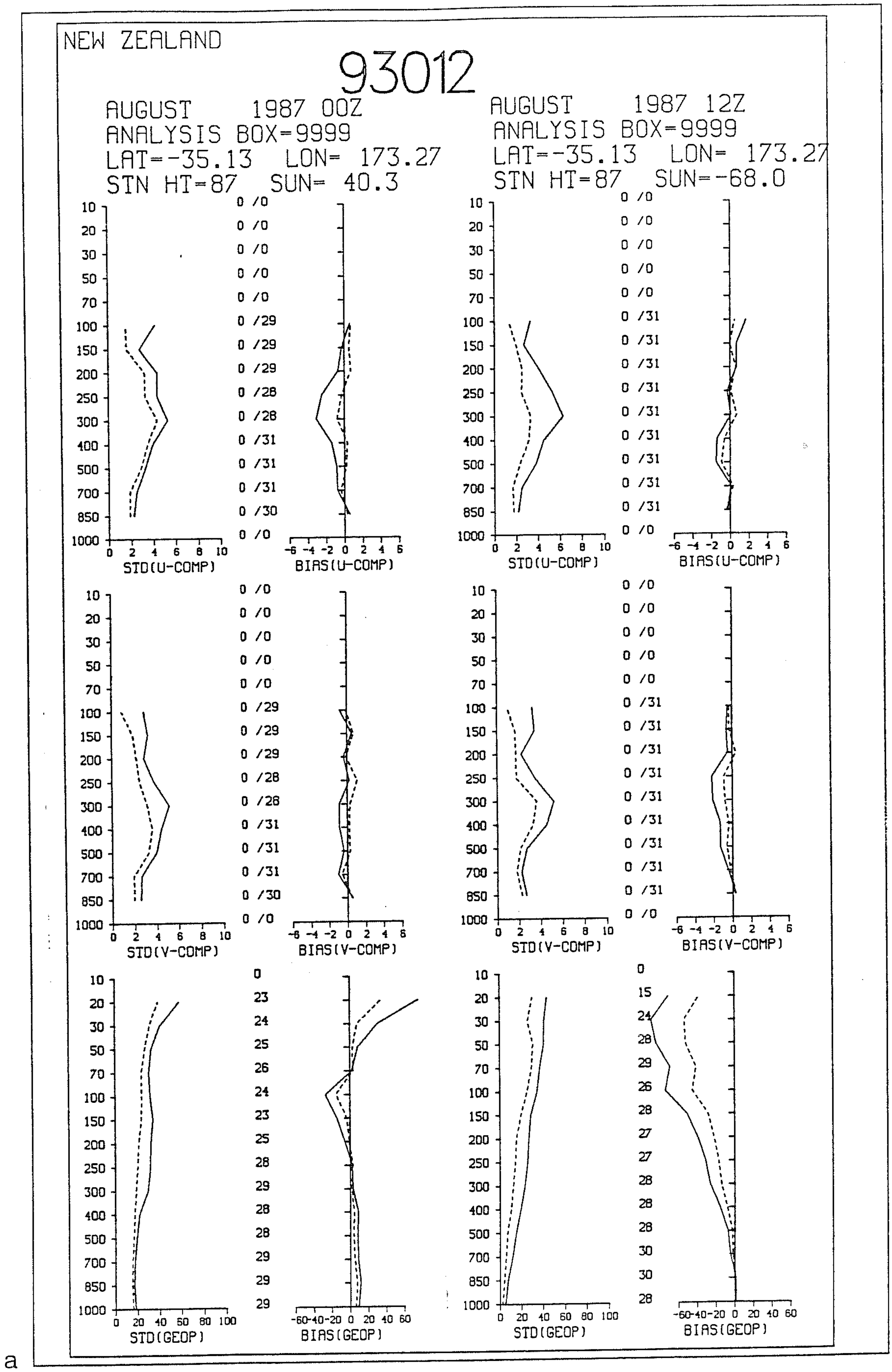
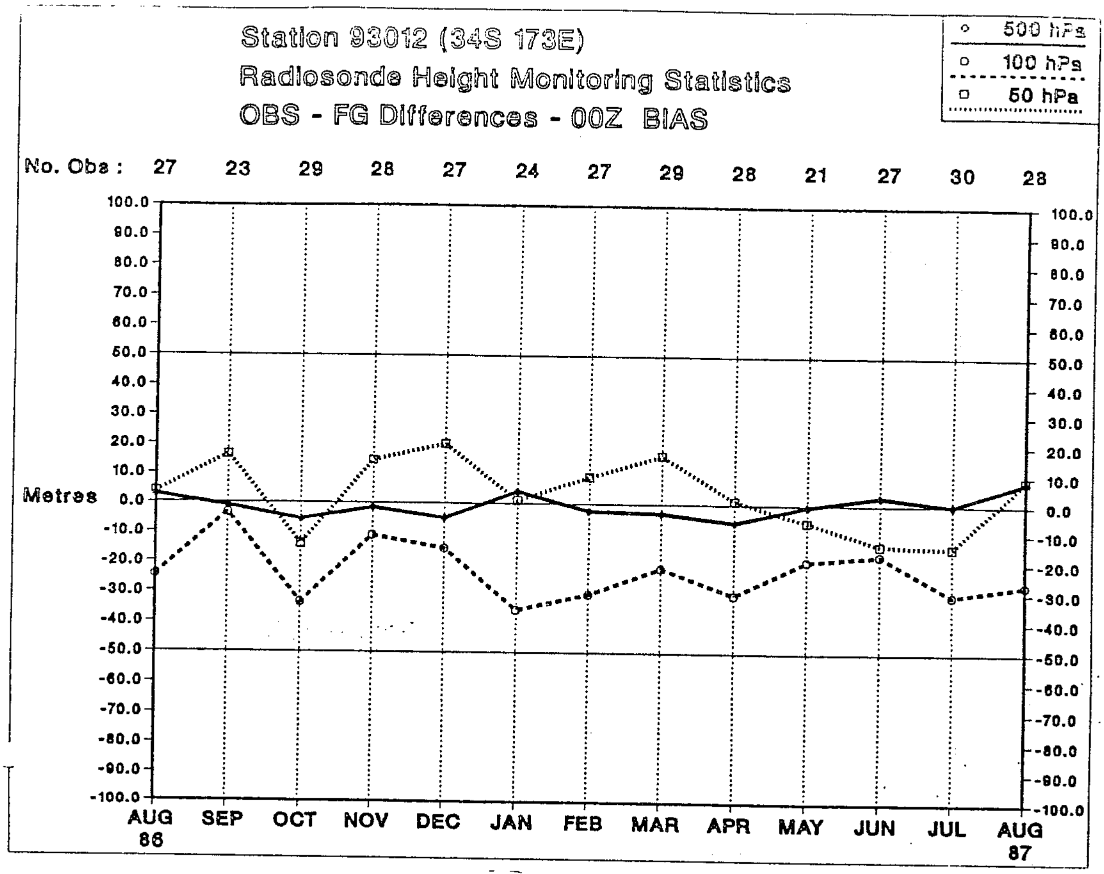
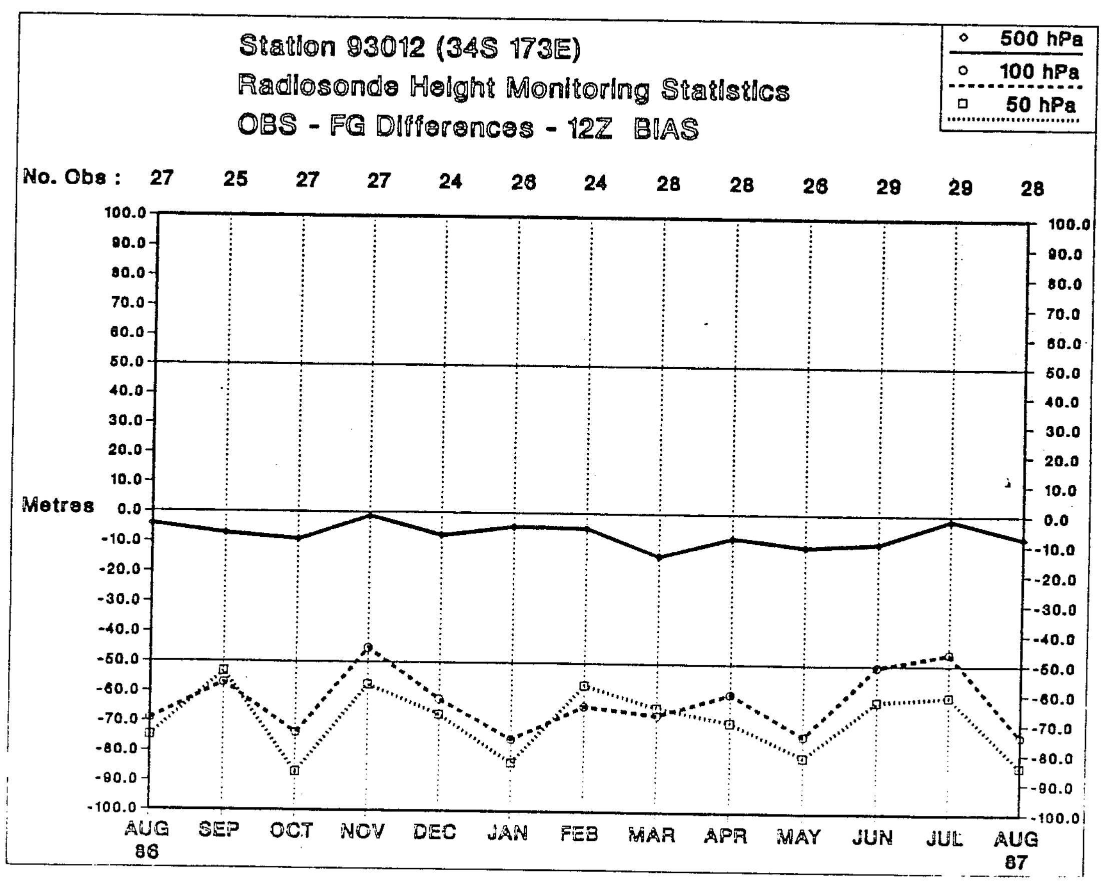


Fig. 1. For station 93012, Kaitaia Observatory, 35° 08'S, 173° 16'E and August 1987. (a) Differences between height and wind observations and the ECMWF first guess and analysis (dashed lines denote deviations from uninitialised analysis, solid lines deviations from first guess fields). (b) Bias curves, for 0000 and 1200 UTC, showing the evolution of the mean monthly differences between observations and the first guess over a 13 month period.



Station 93012 (34S 173E)
Radiosonde Height Monitoring Statistics
OBS - FG Differences - 00Z BIAS
500 hPa
100 hPa
50 hPa
No. Obs : 27 23 29 28 27 24 27 29 28 21 27 30 28
Metres
AUG 86
SEP
OCT
NOV
DEC
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG 87

b



Station 93012 (34S 173E)
Radiosonde Height Monitoring Statistics
OBS - FG Differences - 12Z BIAS
500 hPa
100 hPa
50 hPa
No. Obs : 27 25 27 27 24 26 24 28 28 26 29 29 28
Metres
AUG 86
SEP
OCT
NOV
DEC
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG 87

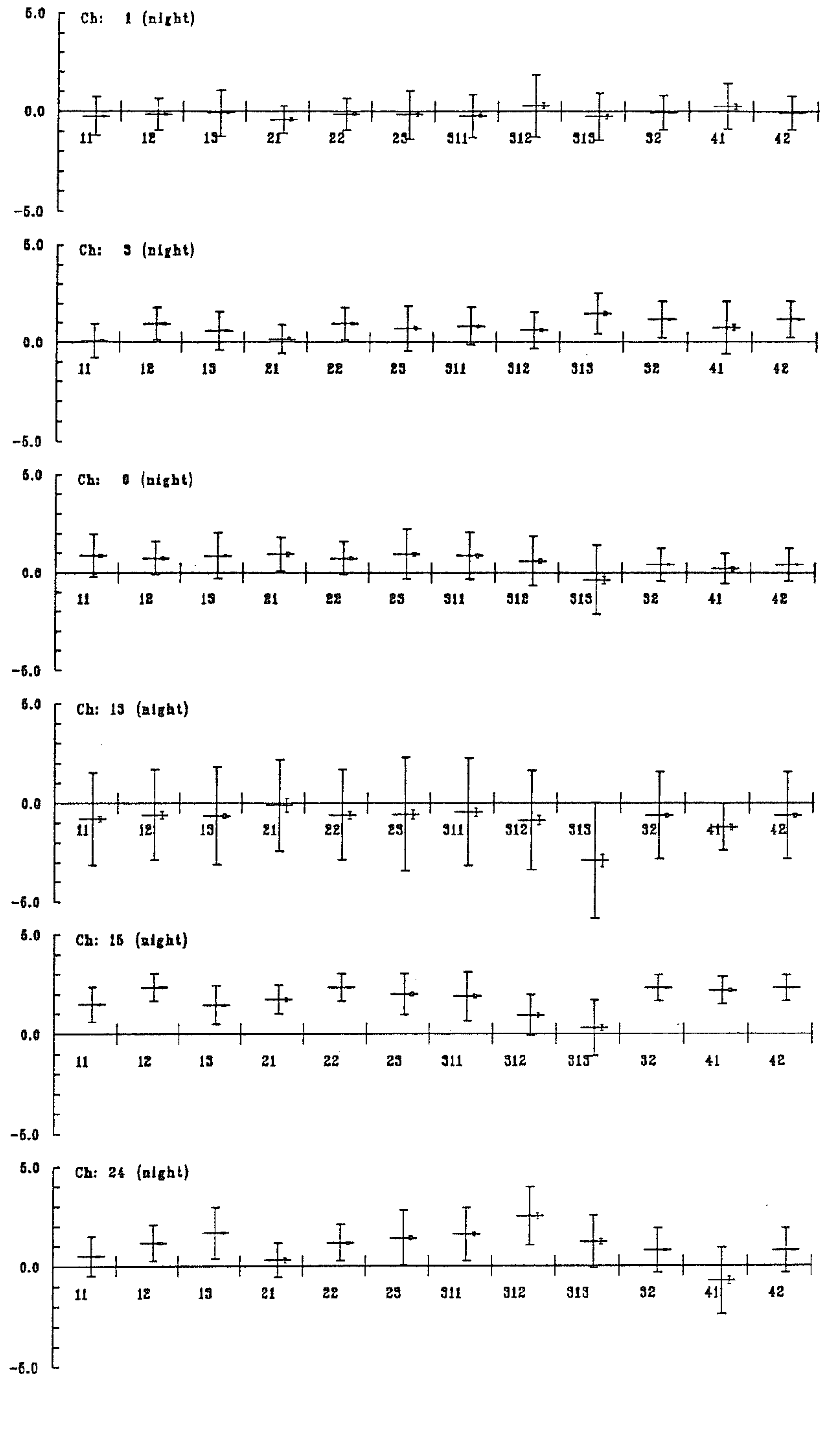
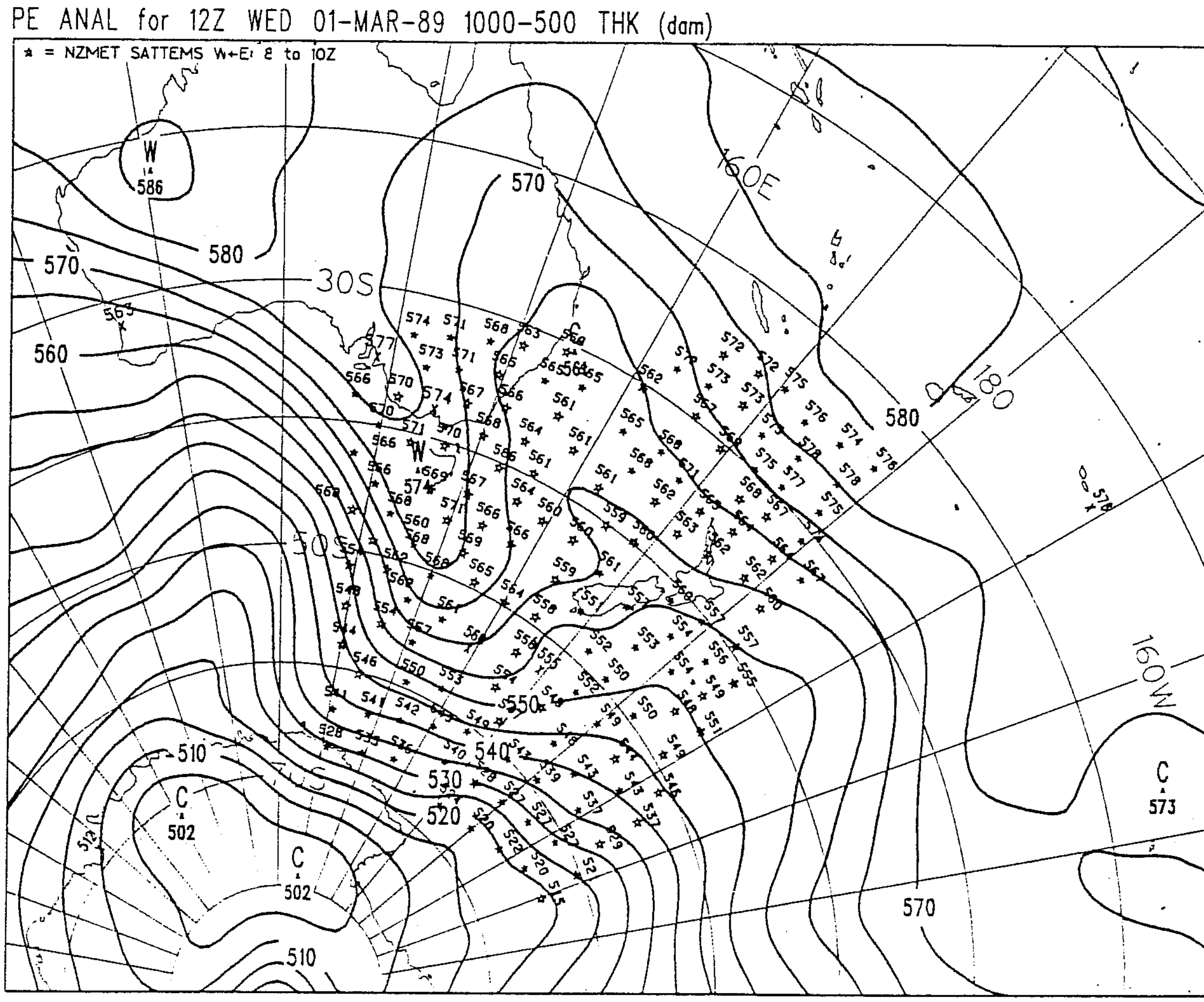


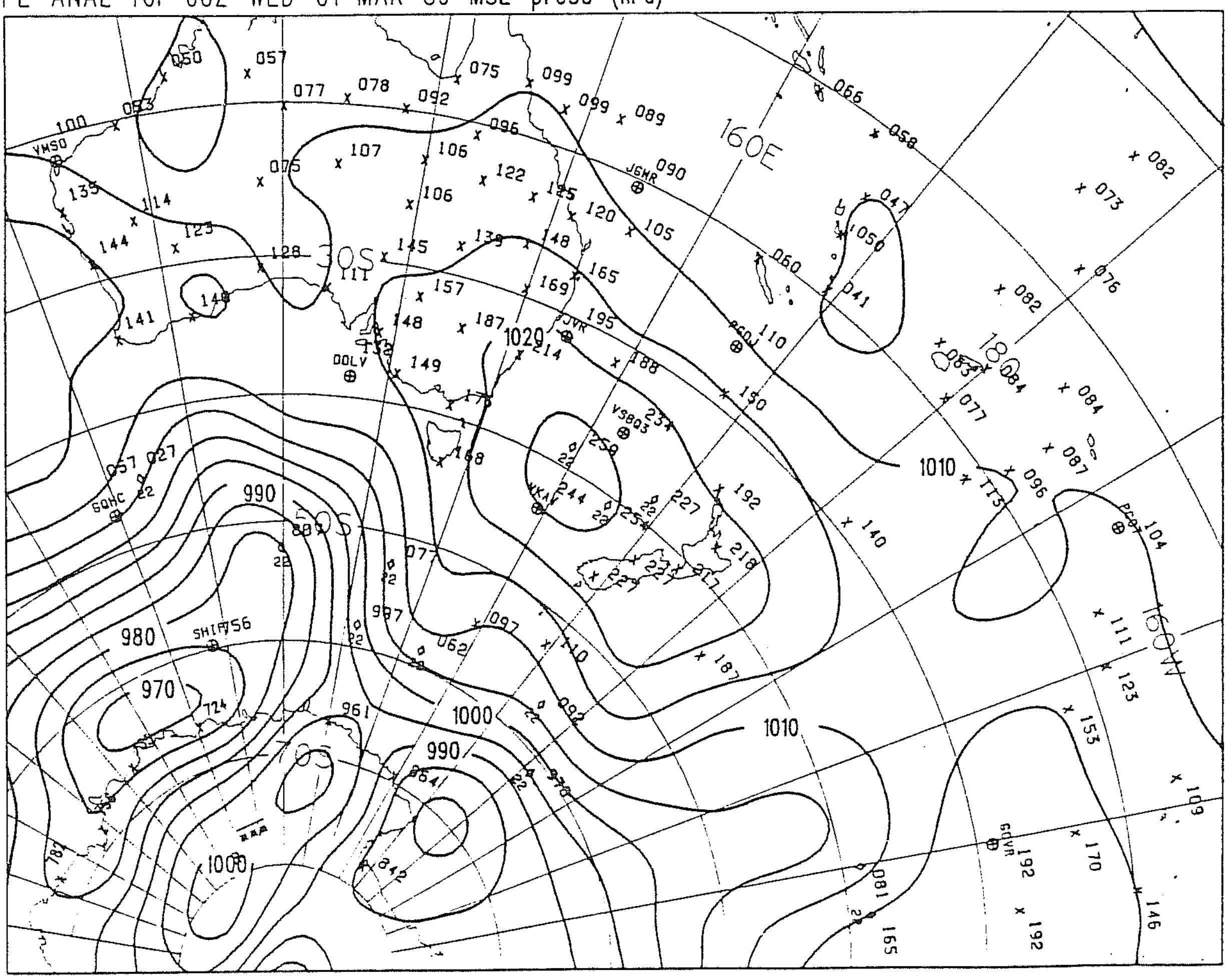
Fig. 2. Mean, standard deviation (large error bars) and standard error (small error bars) statistics for a selection of NOAA10 channel δ corrections, plotted as a function of an air-mass index (see Uddstrom, 1988b).



a

Fig. 3. Analysis fields for the NZMS regional model grid. (a) 1000 – 500 hPa analysis for 1200 UTC. Radiosonde measured thicknesses are indicated by Xs, while NZMS satellite derived thicknesses are indicated by "⋆" symbols. Open ⋆ s indicate better quality retrievals than do closed ⋆ s. (b) MSL pressure analysis for 0000 UTC. Drifting buoy reports are indicated by "◇" symbols, and ship reports by "⊕" symbols. Land stations are indicated by Xs. An "R" indicates a rejected data item. Note, at this scale, not all data are shown.

PE ANAL for 00Z WED 01-MAR-89 MSL press (hPa)



b

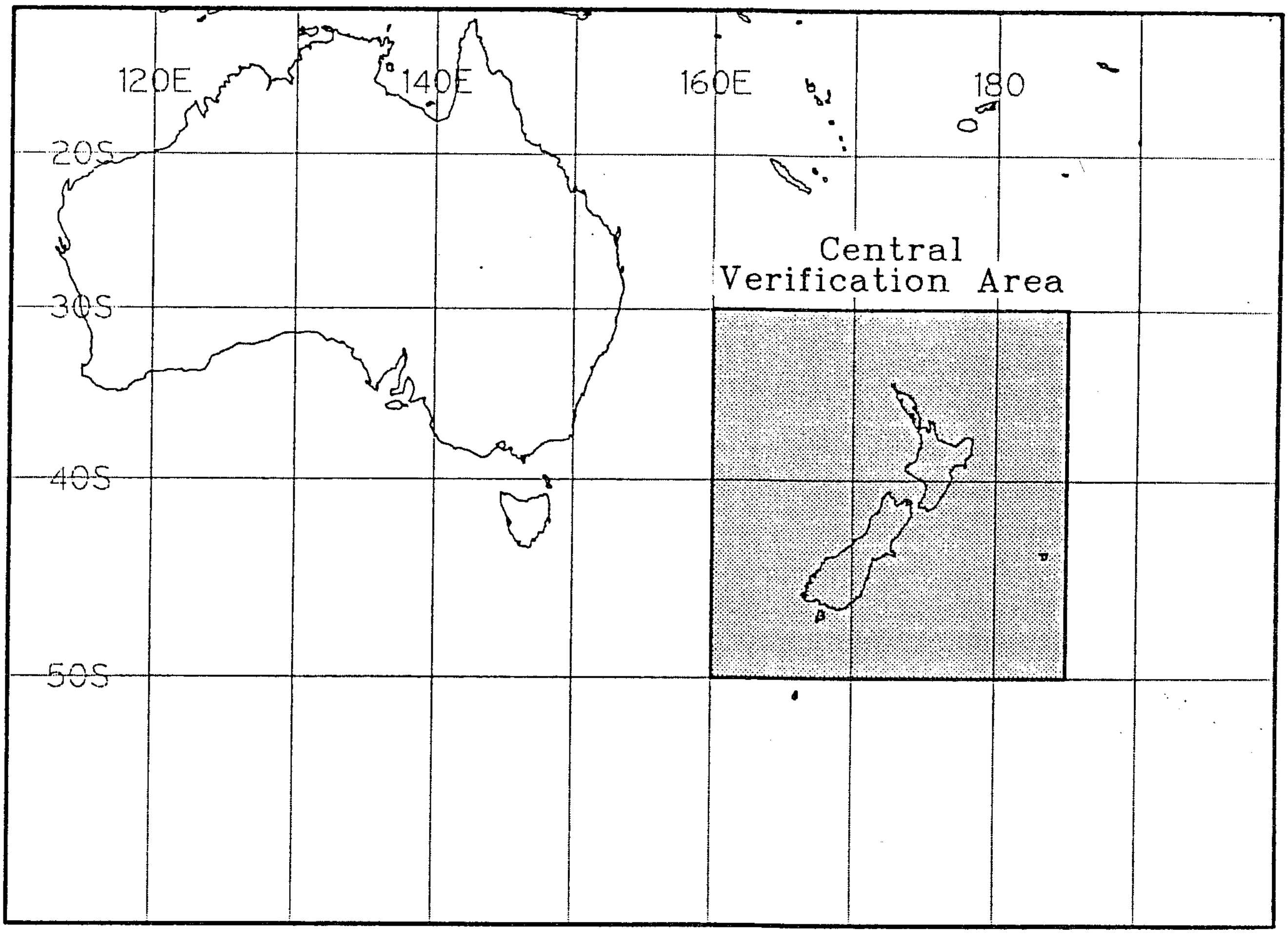


Fig. 4. Area for which verification statistics are computed.

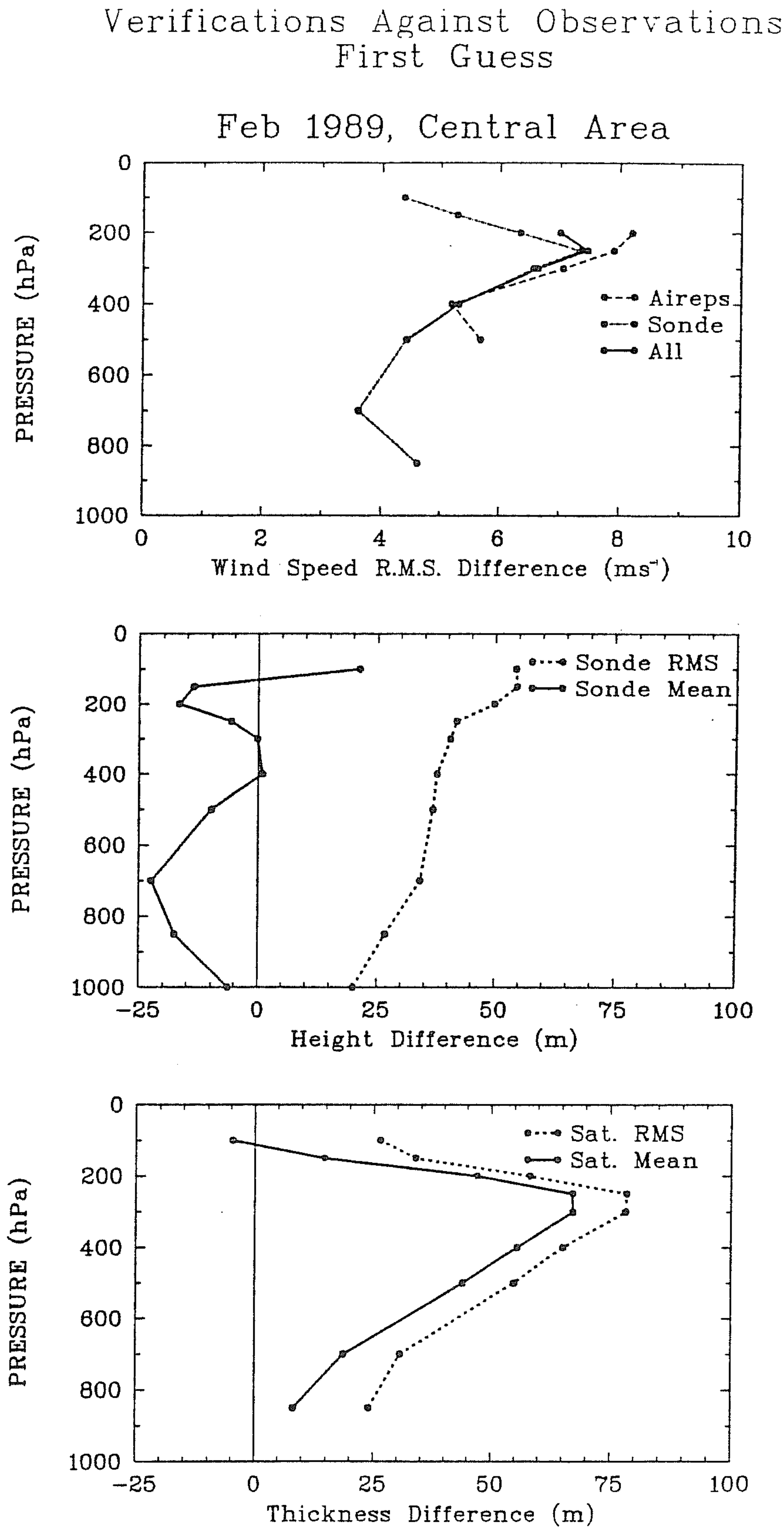
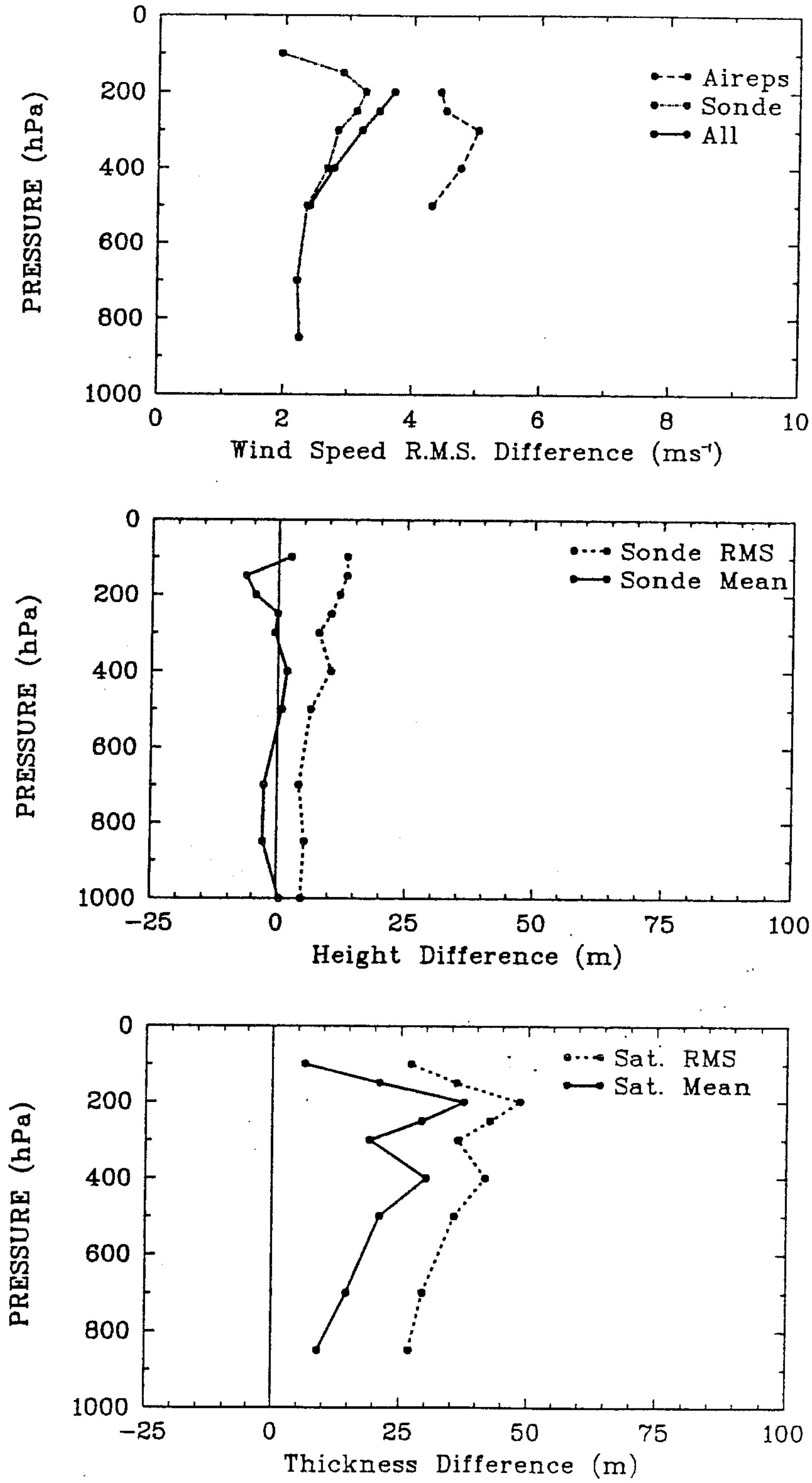


Fig. 5. Assimilation scheme inferred error statistics for data types AIREPS, upper TEMPS, upper WINDS, satellite thicknesses, drifting buoys, SHIP reports, LAND stations, and BOGUS surface pressures. Both observation minus first guess, and observation minus analysis statistics of bias and rms difference are given.

Verifications Against Observations.
Analyses

Feb 1989, Central Area



Verifications Against Observations.
MSL Pressure, Feb 1989
Central Area