DATA PROCESSING, SELECTION AND QUALITY CONTROL AS USED AT JMA

Masao Kanamitsu*

European Centre for Medium Range Weather Forecasts
Reading, U.K.

1. INTRODUCTION

This paper describes the detailed quality control, data checking and data selection procedures used in the Japan Meteorological Agency (JMA) analysis/forecast system as of September 1984. There are several unconventional types of data used at JMA, namely: humidity data obtained from the Geostationary Meteorological Satellite (GMS),cloud data and tropical cyclone bogus data. In addition, statistical corrections of the radiosonde observations in the stratosphere have been used operationally for the last two years. These aspects of the data handlings are also described in detail in this paper.

2. ASSIMILATION SYSTEM

The JMA analysis/forecast system is summarised in Figs. 1a and 1b. Major features of the system to be emphasized in relation to the use of data are:

(a) two dimensional multi-variate analysis of winds and height, but univariate analysis of temperature and dew point depression; surface pressure is analysed by a univariate scheme in the northern hemisphere but by a multivariate scheme in the southern hemisphere.

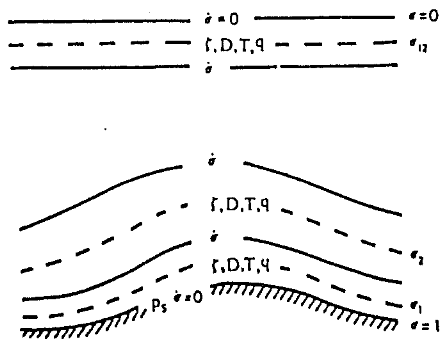
(b) functional fitting analysis scheme in the stratosphere.

(c) forecast guess created by the 12 level global spectral model with a triangular 42 truncation horizontal resolution.

(d) 12 hour intermittent data assimilation.

* on leave from the Japan Meteorological Agency, Tokyo, Japan

ANALYSIS	PREDICTION
z, u, v, T, q	ζ, D, T, q
p(mb)	sigma
10	0.025 (sigma 12)
20	0.075
30	0.160
50	0.290
70	0.430
100	0.565
150	0.690
200	0.800
250	0.890
300	0.950
400	0.980
500	0.995 (sigma 1)
700	
850	q analyzed
1000	below 400mb



Model sigma levels and arrangement of variables. Analyses levels are also indicated at the left side.

Independent variables	λ, ϕ, σ, t
Dependent variables	T, ψ, χ, π, q
Diagnostic variables	Φ, $\dot{\sigma}$
Integration domain	Global
Time integration	Semi-implicit scheme
Earth surface	Albedo, soil-moisture, snow and ice specified geographically.
Orography	Included, with a modification to smooth out small scale features.
Physical Parameterization	(i) horizontal diffusion : Linear, fourth order (ii) vertical diffusion : K-theory dependent on Richardson Number. (iii) Kuo-type convection scheme (iv) radiation, calculated every 3 hours. (v) computed land temperature, diurnal cycle. (vi) climatological sea-surface temperature. (vii) surface fluxes : momentum, sensible heat, moisture

Fig. 1a The JMA forecast model used for data assimilation

Integration domain	Global
Initial time	00Z, 12Z
Cut-off time	6 hours after map time
Data assimilation frequency	12 hour
Analysis first guess	12-hour forecast
Analysis method	
for troposphere	Two-dimensional multi-variate optimum interpolation for geopotential height and wind. Two-dimensional uni-variate optimum interpolation for the others.
for stratosphere	The least square best fitting
Initialization	Non-linear normal mode
Numerical model	12-level T42 global model
Statistical correction of forecast	None
Analysis used for verification	objective analysis

Fig. 1b The JMA analysis system

Further details of the forecast model and the analysis system are described in Kanamitsu et al. (1983), Kashiwagi (1984) and Electronic Computation Center (1983).

3. FLOW DIAGRAM

The flow diagram of the JMA analysis/forecast system is shown in Fig. 2. The data from the Global Telecommunication System (GTS) are first processed by a communications computer named ADESS (Automatic Data Editing and Switching System), and are then sent to the Front End Processor (FEP) which is connected to the main analysis/forecast system computer (HITAC M-200H). The data stored in the FEP disks are then ready to be accessed from the main computer system. No data checking is performed up to this point.

The first step in the analysis system is the decoding of reports and the creation of a data file suitable for input to the analysis programme. The second step is to geographically sort the data. At this stage, the first major data checking and quality control are performed. These are mainly composed of a vertical consistency check of TEMP and SATEM data. The processing of the GMS cloud data for humidity and the statistical correction of TEMPS in the stratosphere is performed between the first and the second step. The data are then passed to the analysis step and the final horizontal checking and merging of closely located data are performed. The rejected and corrected data are archived for monitoring purposes.

Further details about the computer system and the flow diagram can be found in Electronic Computation Center (1983).

4. TREATMENT OF DATA

4.1 Data types

The types of data and the variables used in the analysis/forecast system are listed below:

```
                     |
                     | (Input)     ADESS                    | (Output)
                     |                                      |
             +-------+-------+                      +-------+-------+
             | Communication |                      | Communication |
             |      and      |                      |      and      |
             |    Sorting    |                      |    editing    |
             +-------+-------+                      +-------+-------+
   +---------+       |                                      |
   | Typhoon |       |                                      |
   |  bogus  +-------+                                      |
   +---------+       |                                      |
             +-------+-------+                      +-------+-------+
        +---+    Decoding    +---+                  |     Coding    |
        |   +-------+-------+    |                  +-------+-------+
  +-----+-----+     |        +------+-----+                 |
  |    GMS    |     |        |  Strato.   |                 |
  |  humidiy  |     |        |   TEMP     |                 |
  | processing|     |        | correction |                 |
  +-----+-----+     |        +------+-----+                 |
        |   +-------+-------+       |                       |
        |   |    Sorting    |       |                       |
        +---+      and      +---+                           |
            |  vert. check  |                               |
            +-------+-------+                               |
                    |                                       |
            +-------+-------+                               |
            | Horiz. check  |                               |
        +- -+      and      +---+                           |
        |   |  OI analysis  |   |                           |
        |   +-------+-------+   | (12 hour forecast)        |
        |           |           |                           |
        |   +-------+-------+   |                           |
        +   +    Forecast   +---+                           |
        |   +-------+-------+                               |
        +-----------+                                       |
                    |                                       |
        +-----------+-----------+                           |
        |                       |                           |
  +-----+-------+         +-----+----+                      |
  | Application |         | Graphics |                      |
  +-----+-------+         +-----+----+                      |
        |                       |                           |
        +-----------------------+---------------------------+
```

Fig. 2 Flow diagram of the JMA analysis/forecast system.

(a) TEMP-A, TEMP-B, TEMP-C, TEMP-D ($\underline{V}$, z, T, Td)

(b) SATEM-A, SATEM-B (Thickness)

(c) PILOT ($\underline{V}$)

(d) SATOB ($\underline{V}$)

(e) SYNOP (p, $\underline{V}$, T, Td)

(f) SHIP (p, $\underline{V}$, t, Td)

(g) AIREP ($\underline{V}$)

(h) GMS (Cloud data)

(i) BOGUS (p, $\underline{V}$, z)

(j) AMEDAS (Rain)

The GMS and BOGUS data are non-conventional. The AMEDAS (Automatic Meteorological Data Acquisition System) is a regional high density automatic surface observation over Japan which is currently used for verification purposes and therefore further details will not be given in this paper.

4.2 Cut-off time

The data cut-off time is 6 hours after the observation (both for 00Z and 12Z). The data within plus or minus 6 hours of analysis time are used. Observational errors are increased according to the time difference between the observation and analysis.

4.3 Use of the data in the analysis

TEMPS and PILOTS
Standard pressure level data are used in the analysis; significant level data are only used for checking the vertical consistency of soundings.

The observation errors of the TEMPS are modified for some stations based on the statistics of the fit of the observations to the initial guess field.

The solar correction of the stratospheric soundings (above 100 mb) is performed using a simple statistical correction of the day soundings to the night soundings. Details of the correction procedure are described in Appendix 1.

AIREPS

The data are assigned to the nearest analysis level. Only wind data are used.

SATOBS

The data between 900 and 650 mb are assigned to the 850 mb level, whilst data between 350 and 70 mb is used for the analysis of the 300, 250, 200 and 150 mb levels. The observational errors are allowed to vary with height based on the SATOB-TEMP collocation statistics.

SATEMS

The thickness is converted to the temperature at the analysis levels using a cubic spline; the interpolated temperature is used to compute the thickness between analysis levels. The thickness is then added to the analysed height at the level immediately below the level to be analysed so as to form height data. Note that the surface pressure is first analysed and used as the first reference level. This method effectively reduces the bias error of the initial forecast guess field.

The major weakness of this method is that the reference level is always chosen from the level below; therefore no information from higher levels is used. The use of 3-dimensional OI seems to be the most natural approach to remove this weakness.

GMS data

The GMS data provided by a geostationary satellite consists of cloud cover, mean cloud top temperature, spatial variance of cloud top temperature and maximum/minimum temperature over a one degree latitude/longitude grid. They are computed from picture element (pixel) data whose typical resolution is about 5 km. These data are used to generate moisture soundings over the area covered by the GMS (approximately 90^{o}E - 170^{o}W, 50^{o}S - 50^{o}N). The detailed procedure is presented in Appendix 2 along with the accuracies of the derived soundings. The impact of the use of the data on the forecasts have been found to be significant particularly for convectively driven systems in the summer season.

BOGUS

Typhoon bogus data and PAOBS (bogus surface data created by the Australian Bureau of Meteorology) are the major data in this category. Details about the typhoon bogus data are given in Appendix 3.

All the bogus data are regarded as very accurate and observation errors which are less than those for TEMP data are assigned to them.

5. DATA CHECKING AND QUALITY CONTROL

Table 1 shows a summary of the data checking and quality control procedures applied in the JMA system. In the following paragraphs, each procedure is described in detail.

5.1 Decoding

The first quality control of data is performed in the decoding process. Several complicated steps are programmed to recover some of the apparent coding errors due to transmission and simple human error. The items examined are the heading, date, identification of data, number of characters and the key words. If the errors are found to be unrecoverable, the data are discarded. In order to remove apparently erroneous data resulting from such a mechanical recovery process, checks of absolute magnitudes are also made. Proper reliabilities are given to the recovered data.

Duplicate data are removed by giving the highest priority to the incoming trunk circuit. This is effective in maintaining the good quality of observations which are normally transmitted through the major trunk circuit. Correction reports are used to replace the original data.

Table 1 Quality control and data checking of JMA system

Process	Checking item	Action taken
Decoding	Header	Delete
	Key word	Delete
	Number of characters	Correct
	Absolute magnitude	Delete
Sorting	Duplicate	Priority placed
	Correction report	Priority placed
Closely located data	Distance between the stations for SYNOP, SHIP, DRIBU, AIREP, SATOB.	If less than 50km, one datum chosen.
	Distance between the SATEM observations.	If less than 200km, one datum chosen.
	Distance between the TEMP and SATEM observations.	If less than 200km, SATEM data are discarded.
Vertical check (TEMP)	Absolute magnitude	Delete
	Temperature lapse rate	Delete/correct
	Consistency between sig. and std. pressure level data	Delete/correct
	Magnitude of wind shear	Delete/correct
	Hydrostatic relation	Delete/correct
	Icing	Delete
	Lowest level pressure against station altitude.	Delete
Vertical check (SATEM)	Reliability	Delete
	Temperature lapse rate	Delete
Horizontal check (tropo-Sphere)	Check against first guess	Delete
	Check against neighboring stations.	Delete
Horizontal check (strato-sphere)	Check against persistence	Delete
	Check against preliminary analysis.	Delete

Some statistics of the relation between the rejected data in the assimilation cycle and the data corrected in the decoding process have been collected. It is found that the data corrected in the decoding process have a larger rate of rejection, indicating the lower quality of such observations (refer to Fig. 3 and Table 2).

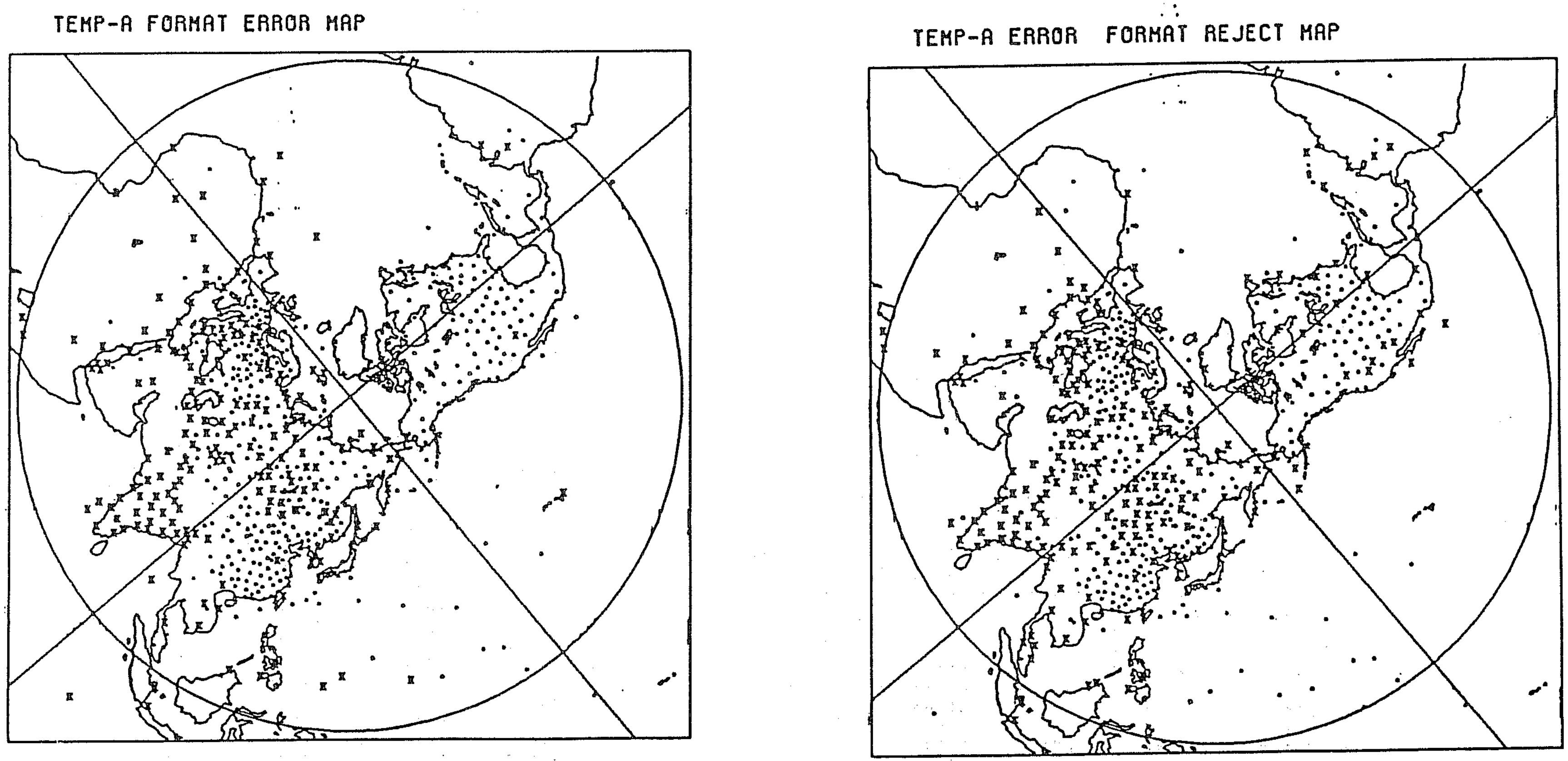
5.2 Vertical check

The major data checking is performed in the sorting step. The checks are mainly made for the vertical consistency of a sounding. The main purpose of this check is to maintain the high quality of TEMP data. Since observational errors of TEMPS are assumed to be small compared to most other observing systems, TEMP observations are normally given a large weight in the analysis. Therefore, it is crucial to control the quality of TEMP data.

TEMP data
Major items checked for this type of data are listed below:

(a) absolute magnitude
(b) magnitude of vertical wind shear
(c) temperature lapse rate
(d) consistency between significant and mandatory level data
(e) hydrostatic relation
(f) lowest level pressure against station altitude
(g) instrument icing.

For (a) and (b), histograms and variances are used to determine threshold values for the detection of erroneous data. For the temperature lapse rate check (c), the static stability of the soundings is first examined. If the stratification is found to be dry unstable, then the sounding is adjusted to the neutral state (using a dry convective adjustment method) and the deviation of the observed profile from the adjusted one is used as a measure of the quality of the data. Histograms and variances are computed for the deviations and the threshold valued are determined.



(a) (b)

Fig. 3 Maps of the TEMP stations (marked 'X') with above average number of format errors (a), and above average rejection by the analysis system (b). Good correspondense between (a) and (b) implies poor quality of corrected data in the decoding process.

Table 2 Rejection rate of the reports corrected and uncorrected in the decoding step.

P	SPEC	NFT	NFR	EFT	EFR
850	Z	99.23	0.77	77.26	22.74
	T	97.83	2.17	90.81	9.19
	TD	99.61	0.39	98.26	1.74
	V	99.40	0.60	98.00	2.00
700	Z	99.30	0.70	93.23	6.77
	T	98.93	1.07	89.53	10.47
	TD	99.36	0.64	98.06	1.94
	V	99.74	0.26	99.06	0.94
500	Z	99.01	0.99	89.37	10.63
	T	98.96	1.04	86.29	13.71
	TD	99.33	0.67	98.46	1.52
	V	99.64	0.36	98.19	1.81
400	Z	98.87	1.13	90.11	9.89
	T	98.84	1.16	88.12	11.88
	TD	99.59	0.41	97.42	2.58
	V	99.28	0.72	97.32	2.68
300	Z	98.38	1.62	86.76	13.24
	T	98.30	1.70	85.58	14.42
	TD	99.77	0.23	96.48	3.52
	V	98.92	1.08	97.02	2.98
250	Z	98.10	1.90	87.04	12.96
	T	97.54	2.46	84.28	15.72
	TD	99.80	0.20	96.88	3.13
	V	99.09	0.91	99.11	0.89
200	Z	97.51	2.49	84.52	15.48
	T	97.33	2.67	84.93	15.07
	TD	99.86	0.14	96.95	3.05
	V	99.50	0.50	98.21	1.79
150	Z	96.82	3.18	80.92	19.08
	T	97.96	2.04	84.72	15.28
	TD	99.81	0.19	96.47	3.53
	V	99.59	0.41	97.54	2.46
100	Z	95.82	4.18	83.70	16.30
	T	97.22	2.78	86.49	13.51
	TD	99.78	0.22	100.00	0.00
	V	99.67	0.33	99.53	0.47

P	SPEC	NFT	NFR	EFT	EFR
70	Z	97.62	2.38	84.44	15.56
	T	98.33	1.67	58.60	41.40
	TD	-	-	-	-
	V	99.59	0.41	98.85	1.15
50	Z	97.43	2.57	81.63	18.37
	T	98.81	1.19	90.05	9.95
	TD	-	-	-	-
	V	99.80	0.20	98.99	1.01
30	Z	96.65	3.35	86.89	13.11
	T	98.67	1.33	90.67	9.33
	TD	-	-	-	-
	V	99.39	0.61	98.78	1.22
20	Z	97.31	2.69	80.21	19.79
	T	98.44	1.56	88.89	11.11
	TD	-	-	-	-
	V	99.60	0.40	96.12	3.88
10	Z	95.34	4.66	62.07	13.24
	T	97.63	2.37	90.32	14.42
	TD	-	-	-	-
	V	96.22	3.78	-	-

* NFT ; NORMAL FORTMAT THROUGH RATE (%)
* NFR ; NORMAL FORTMAT REJECT "
* EFT ; ERROR FORTMAT THROUGH "
* EFR ; ERROR FORTMAT REJECT "

NFT = (T.T+M.T)/(T+M+R) FOR NORMAL FORMAT
NFR = (R+T.R+M.R)/(T+M+R) "

EFT = (T.T+M.T)/(T+M+R) FOR ERRO FORMAT
EFR = (R+T.R+M.R)/(T+M+R) "

The consistency between the significant and mandatory level data (d) is checked by using deviations of mandatory level data from linearly interpolated values derived from the data at the two adjacent significant levels. Similarly, a hydrostatic check (e) is performed using the difference between the thickness obtained from the standard pressure level heights and the thickness computed from the reported temperature and humidity. Examples of the histograms are presented in Fig. 4.

In the first step of the vertical check, erroneous reports are itentified using the threshold values mentioned above and by applying the condition that smooth vertical soundings are obtained with minimum deletion or correction of reports in (b),(c) and (d). In the case of the hydrostatic check (e), the distinction between the directly observed and derived quantities is also taken into consideration.

The second step of the vertical check is to correct or fill-in the erroneous or missing reports by vertical interpolation. In order to maintain the quality of TEMP data, only height and temperature data are corrected or filed.

The check of the lowest level TEMP data (f) is made using station altitude. If the data is found to be erroneous, all the reports based on the lowest level are discarded.

The instruments icing (g) is checked by looking at the vertical profile and the data are discarded if icing is suspected.

SATEMS

For SATEM data, a vertical check of the temperature lapse rate against dry unstable stratification is performed, but no attempt is made to correct the erroneous data.

5.3 Horizontal check

The final data checking is performed in the analysis program. In the

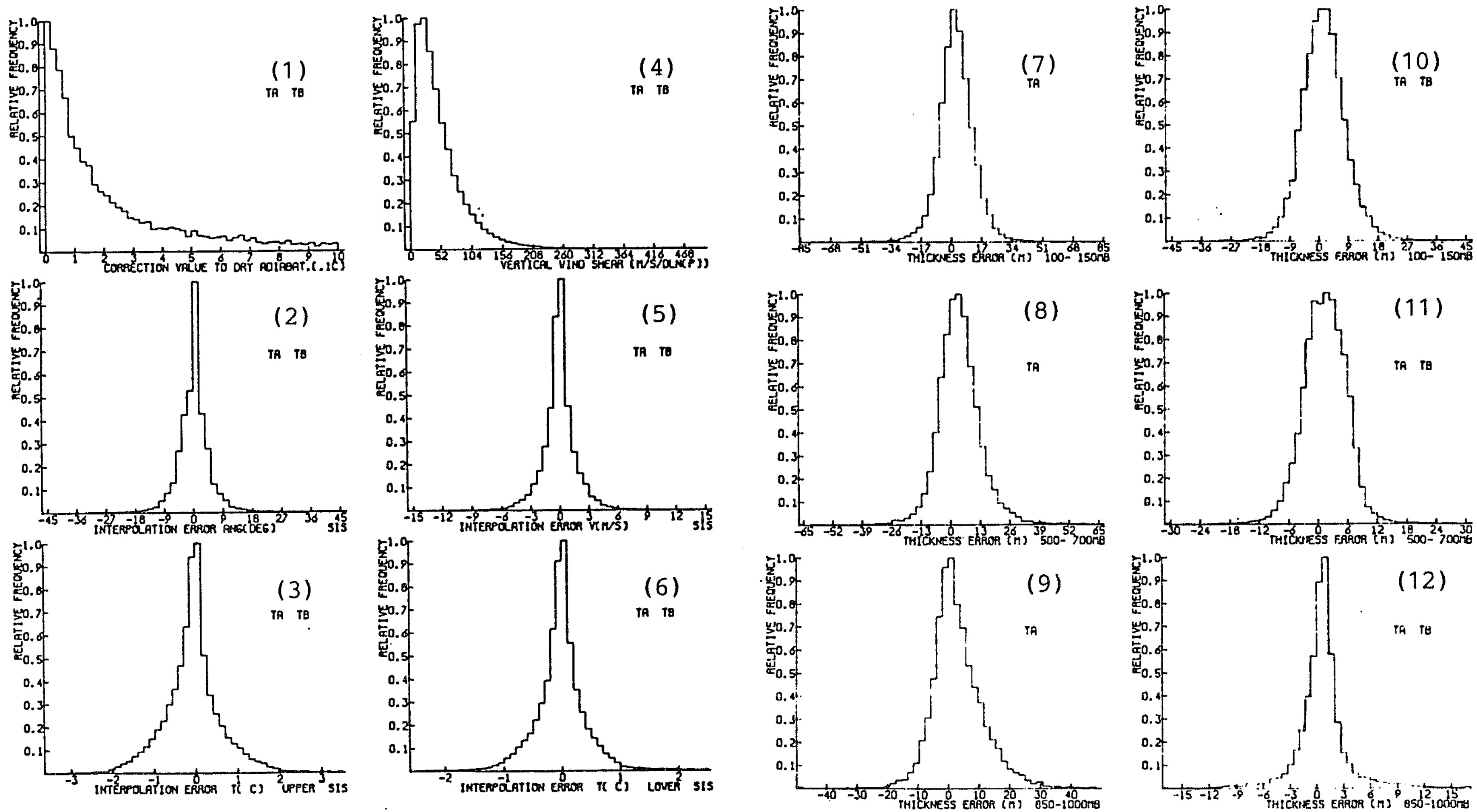


Fig. 4 Histograms used for the vertical consistency check of TEMP. See text for the detailes of the adjusted, interpolated and computed soundings.

(1) Deviation of temperature from the adjusted soundings.
(2) Deviation of wind direction from the interpolated soundings.
(3) Deviation of temperature from the interpolated soundings (100-300mb).
(4) Vertical wind shear.
(5) Deviation of wind speed from the interpolated.
(6) Deviation of temperature from interpolated (400-1000mb).
(7) Deviation of thickness from computed soundings (100-150mb).
(8) Deviation of thickness from computed soundings (500-700mb).
(9) Deviation of thickness from computed soundings (850-1000mb).
(10) Same as (7) but computed with significant level data.
(11) Same as (8) but computed with significant level data.
(12) Same as (9) but computed with significant level data.

troposphere between 1000 and 100 mb, the data are checked against first guess fields.

- If the deviation from the guess is greater than a prescribed value (e_1), then the data is discarded.
- If the deviation is less than e_1 but larger than another prescribed value e_2, then the check against neighbouring observations is performed. In the buddy check, the successive correction method is used to interpolate the neighbouring observations to the observation point to be checked.
- If the difference is greater than e_3, then the data is rejected.

The value of e_1, e_2, and e_3 are determined from the root mean square (RMS) fit statistics of observations and forecasts. The values used at present are:

e_1 = 5*RMS

e_2 = 3*RMS

e_3 = given constants.

Modifications are in progress to make e_1 and e_2 functions of the prediction error and e_3 a function of the analysis error.

In the stratosphere, the data are checked against persistence and the preliminary analysis. Note that the curve fitting analysis scheme is used to depict very large horizontal scales of motion in the stratosphere.

5.4 Merging of densely located data

The major purpose of the merging of densely located data is to avoid computational instability in the optimum interpolation scheme. For the data whose observational error is not correlated in the horizontal, namely SYNOP, SHIP, DRIBU, AIREP and SATOB, the data within 50 km are merged into one observation by selecting the data closest to the mean of the closely located data (the reliability and observation time are also taken into account). For SATEMS, whose observational errors are assumed to be horizontally correlated, the distance of the search is increased to 200 km.

For the case of a mixture of SATEMS and other types of data, SATEMS within 200 km of other types of data are discarded. This is based on the assumption that the SATEMS tend to degrade the high accuracy data, such as TEMP observations, when they coexist.

6. DATA SELECTION

The data selection is designed to satisfy the following conditions:

(i) no large discontinuities between grid points,

(ii) weak dependency of the analysis on data selection,

(iii) reasonable computing time

Idealized analysis experiments have been performed to determine the number and types of data to be used in the uni- and multi-variate analyses.

For the uni-variate case, selection of the nearest 10 stations, with the largest weight placed on the data with the smaller observational error, is found to be sufficient.

For the multi-variate case, a simple selection rule is very difficult to find due to the correlations between the observations. For example, the zero correlation between a grid point to be analysed (G) and an observation (A) does not necessarily imply zero weighting of the observation because an additional observation (B) that correlates both with G and A results in non-zero weighting of the observation A. This suggests that a rather large number of observations are required for the multi-variate case to satisfy condition (ii).

With regard to the computer time required, a number of test analyses was subjectively examined and the number of reports was then set to 30 at the lowest level and 45 at higher levels. Also, in order to make analysis increment dynamically balanced, if wind and mass observations are available at one observation point (e.g. TEMPS), all the variables are selected without exception.

The data search starts from a box with 5° latitude/longitude sides centred on the grid point to be analysed; the side of the box is then expanded in steps of 5° until it reaches 30°. If there is one observation or less in the largest box, no analysis is made and the first guess is used.

Acknowledgments

I would like to thank the members of the analysis group of the Electronic Computation Center, JMA for providing material for this paper, and in particular Dr. T. Hiraki who assembled some of the information described.

A number of people have contributed to the work presented here. Mr. K. Kashiwagi designed the tropospheric OI analysis system, Dr. T. Hiraki is responsible for the stratospheric analysis, Mr. N. Miura elaborated the decoding and Mr. T. Nakayama worked on the typhoon bogus.

The procedure to obtain moisture profiles from the GMS cloud data was developed by Mr. S. Isa, and Appendix 2 of this paper is based entirely on his unpublished work.

References

Electronic Computation Center, 1983: Outline of operational numerical weather prediction at Japan Meteorological Agency. Electronic Computation Center, JMA, Tokyo, Japan.

Frank, W.M., 1977: The structue and energetics of the tropical cyclone. I. Storm structure. Mon. Wea. Rev., 105, 1119-1135.

Kanamitsu, M., K. Tada, T. Kudo, N. Sato and S. Isa, 1983: Description of the JMA operational spectral model. J. Meteor. Soc. Japan, 61, 812-828.

Kashiwagi, K., 1984: On the impact of space-based observing system in the JMA global forecast/analysis system. Seminar, 1984. ECMWF, Shinfield Park, Reading, U.K.

Reed, R.J., M.J. Oard and M. Seminski, 1969: A comparison of observed and theoretical diurnal tidal motions between 30 and 60 kilometers. Mon. Wea. Rev., 97, 456-459.

APPENDIX 1

STATISTICAL CORRECTION OF TEMP DATA IN THE STRATOSPHERE

The radiosonde temperature measurements in the stratosphere are known to be affected by the solar radiation. However, the removal of the effect (solar correction) has not been standardized and, therefore, the stratospheric observations are composed of a mixture of processed and unprocessed data.

One of the major problems found at JMA using uncorrected stratospheric data is an unrealistically large diurnal variation of the analysed phase and amplitude of very large scale motions in the stratosphere. In order to eliminate such undesirable variations, we have taken a very simple approach, namely to adjust the day sounding to the night soundings by simply subtracting the day-night differences computed from yearly observations at each station.

Fig. 5 shows the phase-amplitude dial of analysed spherical harmonics (m=1, n=1) of height for uncorrected and corrected data. In the lower panel of Fig. 5, it is clearly seen that the large diurnal variation is removed by this simple method. Table 3 shows the mean correction (day-night difference) of height at 100, 50 and 30 mb for various solar angles over the U.S. The correction for 1982 and 1983 is shown with the corrections reported by Finger (1978) and by Spackman (1965, 1978). There is reasonable agreement between the Finger/Spackman results and ours.

The limitation of the current method is the large variation of the correction from year to year. This implies that either we need more samples to determine corrections or need more frequent updating. It should be further noted that since diurnal and semi-diurnal variations do exist in the stratosphere (Reed et al., 1969) our method may not be entirely appropriate and further improvement is necessary.

The impact of the correction on the stratospheric forecast is found to be significant in several cases.

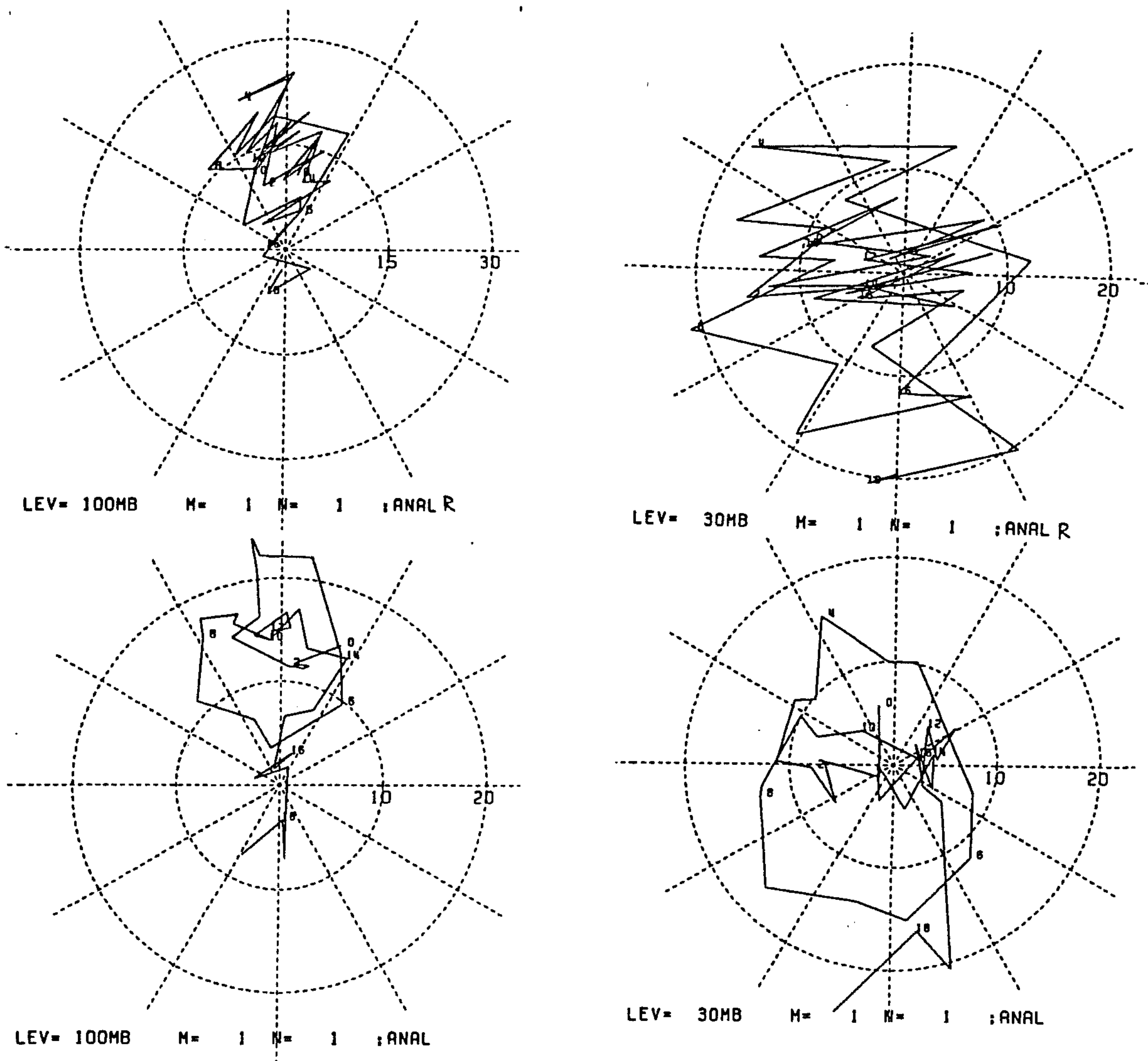


Fig. 5 Phase-amplitude dial of the spherical harmonic component (m=1, n=1) of the analysed height at 100mb (figures on the left) and at 30mb (figures on the right). Top figures are for analyses using uncorrected data and bottom for corrected data.

Table 3 Day-night difference correction of height and temperature of TEMP soundings in the stratosphere for various solar angles (S.A.) for the years 1982 and 1983. The corrections by Spackman (1965, 1978) and by Finger (1978) are also presented.

TYPE OF RADIO-SONDE	PRESSURE (MB)	GEOPOTENTIAL HEIGHT (M)				TEMPERATURE (C)				REFERENCE
		S.A.= 0	S.A.=20	S.A.=40	S.A.=60	S.A.= 0	S.A.=20	S.A.=40	S.A.=60	
U.S.A. (MORNING)	100	10	25	35	40	0.3	0.7	0.9	1.0	FINGER 1965
		-3	13	26	31	0.1	0.6	0.9	0.9	FINGER 1978
		-3	22	*	*	*	*	*	*	SPACKMAN 1978
		-11	15	50	70	0.0	1.2	2.0	1.7	1982
		-14	1	78	60	0.2	0.7	1.9	1.9	1983
	50	18	43	57	64	0.5	1.1	1.3	1.4	FINGER 1965
		0	26	46	51	0.2	0.8	1.1	1.2	FINGER 1978
		9	41	*	*	*	*	*	*	SPACKMAN 1978
		-13	34	84	127	0.0	1.4	2.4	3.0	1982
		-10	8	91	90	-0.2	1.0	2.1	2.1	1983
	30	28	64	80	88	0.8	1.7	1.8	1.8	FINGER 1965
		2	38	64	76	0.3	1.1	1.4	1.5	FINGER 1978
		20	76	*	*	*	*	*	*	SPACKMAN 1978
		-16	57	116	166	-0.1	2.1	2.8	3.4	1982
		-22	17	62	126	0.0	1.1	1.6	2.5	1983

APPENDIX 2

ESTIMATION OF HUMIDITY FIELDS FROM GMS OBSERVATIONS

The following three parameters on a one degree latitude/longitude grid are used to estimate vertical profiles of humidities from the GMS observations.

(1) Cloud amount (Nc) defined as a ratio of the number of pixels with the black body temperature (TBB) less than a given threshold value to the total number of pixels covering the grid area. There are about 400 pixels in the one degree grid area. The threshold value is defined as a function of the surface temperature determined from the IR data.

(2) Average cloud top temperature (Tc) computed from the average of the TBB of cloud covered pixels over the grid area.

(3) Standard deviation of TBB (Sc) of the cloud covered pixels over the grid area.

We first define cloud types depending on combinations of the above three parameters and then relate them to collocated TEMP moisture profiles. The following classification of the parameters is made to define the cloud types.

(a) cloud amount (Nc)

1.	Overcast (OV)	$100<Nc<99\%$
2.	Broken (BK)	$70<Nc<99\%$
3.	Scatter (SC)	$30<Nc<70\%$
4.	Fair (FR)	$1<Nc<30\%$
5.	Clear (CL)	$0<Nc<1\%$

(b) Mean TBB (Tc)

The TBB is converted to pressure using a climatological temperature profile, and the nearest standard pressure level (10 levels up to 100 mb) is assigned (Pt).

(c) Standard deviation of TBB (Sc)

Only in the overcast (OV) case is this parameter used.

1.	Stratus (ST)		$0<Sc<5$
2.1	Stratocumulus 1	(SU1)	$5<Sc<10$
2.2	Stratocumulus 2	(SU2)	$10<Sc<15$
2.3	Stratocumulus 3	(SU3)	$15<Sc$
3.	Cumulonimbus	(CB)	$5<Sc$ and $Pt<250$ mb

(d) Cloud free area

1.	Clear 1	1x1 deg. cloud free area
2.	Clear 2	3x3 deg. cloud free area
3.	Clear 3	5x5 deg. cloud free area

The classification of the cloud types is schematically shown in Fig. 6. There are 60 independent cloud types.

The vertical profiles of moisture from collocated TEMPS are accumulated for each of the cloud type from one year of data. Simple averaging of the samples over the area of GMS coverage is made without any further classification of the soundings, i.e. no distinction between the seasons or latitudes.

Examples of moisture distribution for some of the overcast cloud types are given in Fig. 7. The moisture profiles thus obtained seem to be very reasonable. For example, for the ST-type clouds, the humidity profile shows a maximum at the cloud top level. Furthermore, in the case of CB-type clouds, the entire troposphere is very moist with increasing humidity with the cloud top height.

The standard deviation of the dew point depression for each type of cloud is shown in Table 4. The RMS error of the dew point depression for SATEMS (collocated with TEMP) and the prediction error (12 hour) of the JMA model is also shown in Table 4. It is clear that the standard deviation of the GMS estimated dew point depression over cloudy area is generally much smaller than that of SATEMS, and about equal or smaller than the prediction error. It should be emphasized here that the moisture estimated from the GMS cloud data provides reasonably

```
             (Area)
         +--- CL3
         |
 (Nc)    +--- CL2
         |
 CL  ---+--- CL1

 FR  --- [(9 cloud tops)

 SC  --- [(9 cloud tops)

 BK  --- [(9 cloud tops)

 OV  ---+--- ST --- [(9 cloud tops)
         |
         |                +--- CB  --- [(3 cloud tops)
         +--- SU & CB ---+
                          +--- SU1 --- [(6 cloud tops)
               (Sc)       |
                          +--- SU2 --- [(6 cloud tops)
                          |
                          +--- SU3 --- [(6 cloud tops)

                              (Sc & Pt)
```

Fig. 6 Classification of cloud types. The characters in the parenthesis denote the parameters used for classification. See text for detail.

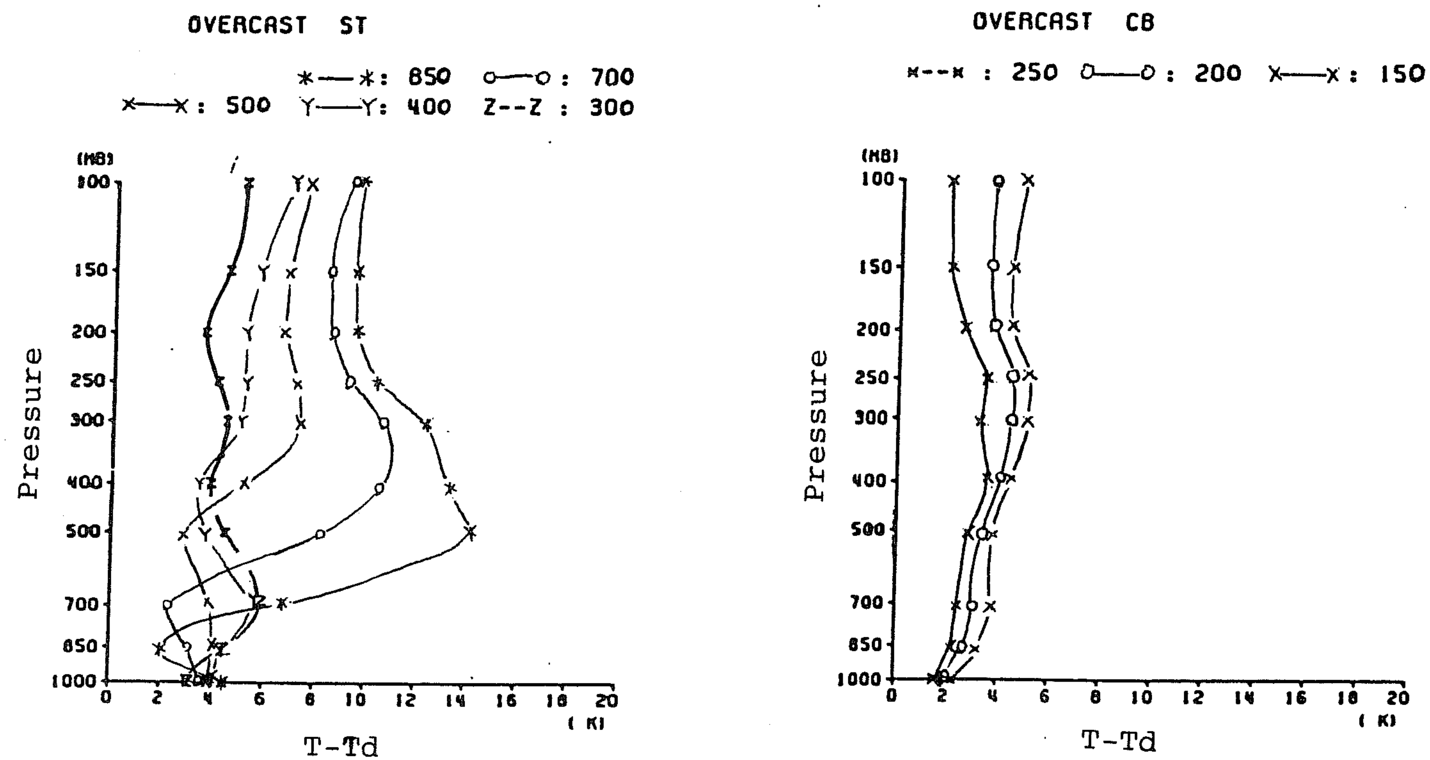


Fig. 7 Examples of the vertical dew point depression profile corresponding to the cloud types. ST-type clouds (left figure) and CB-type clouds (right figure) are shown. The cloud top heights are indicated by various symbols.

Table 4 Standard deviation of the dew point depression colocated with the TEMP for the various types of clouds. The table also includes RMS colocation error of SATEM and the prediction error of the JMA operational spectral model. Unit in deg. K.

Type	850	700	500	400
ST850	2.7	7.4	9.3	7.5
ST700	3.8	3.4	7.7	6.6
ST500	4.9	6.0	3.0	4.3
ST400	5.9	7.5	4.2	2.7
ST300	5.4	7.6	5.2	3.2
CB250	2.5	3.3	3.6	3.7
CB200	2.2	2.4	3.2	3.6
CB150	2.4	2.3	2.3	2.6
CL2	7.2	7.6	6.6	6.1
SATEM	6.4	7.8	8.2	9.6
FCST	5.7	6.3	6.6	6.6

accurate vertical humidity profiles and should be regarded as an important source of moisture data over cloudy areas. It is also noted that the moisture from GMS data has rather a large standard deviation over the clear areas. It is presently planned to use SATEM data over the clear areas.

The use of the GMS cloud data normally produces very different moisture analyses over the tropical and subtropical ocean areas. The impact of the data on the forecast is also substantial in cases where there are convectively driven systems in the tropics.

APPENDIX 3

TYPHOON BOGUS

The objective analyses of tropical cyclones are extremely difficult due to their mixture of large and small scales and to a general lack of observations. However, it is an essential requirement for the operational forecaster to find tropical cyclones in the initial field without significant error of position or intensity for their use in monitoring serious weather events. For this reason, it is necessary to use bogus data for the analysis of tropical cyclones.

The major input for the subjective ananlysis of tropical cyclones is the visible and infrared pictures from the geostationary satellite. Several semi-objective methods are available to determine positions and intensities from such image information. In many cases, reconnaissance flight data from the U.S. Air Force are available in nonconventional form. The subjective analysis of a typhoon based on this information is utilised to define the (1) central position, (2) central pressure and (3) diameter. These parameters are then used to generate idealized typhoon soundings based on the climatology of typhoons (Frank, 1977). About 20-50 bogus vertical soundings of height, temperature and wind around the tropical cyclone are generated. The amount of bogus data depends on the size of the typhoon.

The above procedures are, however, possible only when there are very few actual observations around the tropical cyclone. If the typhoon is close to dense data areas, only purely manual bogussing is performed. It is still not possible to make bogussing in such cases automatic.

The impact of the bogussing on the forecast is extremely large when typhoons are moving into middle latitudes. The intense typhoons are well maintained in the coarse resolution global model such as the one usded at JMA.