Global Criteria for Tracing the Improvements of Radiosondes over the Last Two Decades

Pierre Jeannet, Carl Bower, and Bertrand Calpini

Report for item 6.2 of the

CIMO/OPAG-UPPER-AIR/ET-UASI-2/IOC-2/

28-30. November 2005, Geneva

1. Task and objectives 2
2. Preliminary analysis 3
3. Candidate quality performance criteria 4
4. Geopotential height 7
5. Temperature 9
6. Relative humidity 13
7. Pressure 14
8. Note to the CIMO upper air expert group 15
9. References 16
10. Annex 17

Global Criteria for Tracing the Improvements of Radiosondes over the Last Two Decades

1. Task and objectives

At the last CIMO expert team meeting on upper-air systems intercomparisons (ET on UASI-1, Geneva, 17-20.03.2004), the following task was defined:

Develop performance measures to demonstrate the continuous improvement

in the quality of upper-air observations.

The required action is to "elaborate global criteria for tracing the improvements, based on previous intercomparisons and recent radiosonde development, and including remote sensing". The deliverable will be an "IOM report on global criteria for tracing the improvements of radiosondes"[1]

Here we focus first on the definition of appropriate criteria which should later be used by the CIMO, the National Meteorological and Hydrological Services (NMHS), and the manufacturers in order to trace the improvements in the quality of upper-air observations.

We also apply some of these criteria in order to check performance and present results.

Ideally such criteria should be based on data already available in the previous IOM reports. Guidance on performance criteria are to be included in further CIMO UASI field campaigns as well as national or international intercomparisons to ensure the required continuous tracing.

Finally manufacturers and National Meteorological and Hydrological Services could use these criteria to trace the quality improvements of their radiosondes and sounding stations.

It should be noted that an analysis of the performance improvement has already been prepared by J. Nash and presented at the 2004 Expert team on Upper Air Systems Comparison under the title "Introduction to performance of modern radiosondes based on WMO Radiosonde Comparison results, Temperature". Some of his statements and slides will be used in this report.

[1] A similar task is also defined for surface meteorological measurements. The inclusion of upper-air remote sensing criteria is not addressed in this first issue.

2. Preliminary analysis

The accuracy of radiosondes and the homogeneity of the world upper air network was a matter of concern since already more than 50 years ago, at the time of the first international intercomparisons. International meteorological field campaigns were organized in order to improve the comparability of upper air datasets. The radiosondes manufacturers benefited from these research activities.

Radiosonde sensor technology has definitively improved over the last several decades, but tracing the sensor improvement requires appropriate criteria. Four possible ideas for analyses are suggested by either: (1.) using the previous IOM reports, (2.) comparing radiosonde measurements with model values, (3.) elaborating first a general CIMO questionnaire to the NMHSs, or (4.) extracting numbers from the open literature. They are briefly commented below:

<u>(1). Time evolution of selected statistical parameters reported in the different WMO international radiosonde comparisons</u>.

Apart from the Low-level Intercomparison Experiment, USA, 1979, all intercomparisons with IOM reports took place between 1984 in the UK (phase I), 1985 in the USA (phase II), 1989 in the former USSR (phase III), 1993 in Japan (phase IV), 2001 in Brazil (phase V), and 2005 at Mauritius (phase VI). The statistical parameters (systematic biases, standard deviations, etc) based on differences between the measurements obtained with different types of radiosondes for simultaneous measurements represent a valuable tool for comparison over the last two decades. Each of these campaigns used "link radiosondes" in order to define one reference value (or standard) and have it compared with the other measurements. Nevertheless, Tables 5.3-4 in the Phase II report point to a possible problem with the time continuity of the link radiosondes between Phase I-II-III. At standard levels, we may be able to show better (smaller) differences and perhaps tighter means or RMS values. This analysis has partially been conducted so far: in Table 10, Elms (2003) provides a comparison of the estimated reproducibility of geopotential heights between 500 and 10 hPa for the main radiosonde types obtained from the WMO intercomparisons. Tables 6-8 of the phase IV report provide comparisons between the link radiosondes between phases III and IV.

<u>(2). Time evolution of statistical parameters reported in the different IOM analyses on the comparison between measured and modelled geopotential heights (1988, 1990-1992, 1995-1997, 1998-2001, plus Excel files for the very last years).</u>

In this analysis the comparison is made between radiosonde measurements and the first guess of NWP models, the latter being considered as a relatively stable working reference (Oakley, 1998) although the numerical models underwent significant changes in the last decades. Thus criteria based on the comparison between measurement and model results will inherently have the drawback of adding errors induced by the method of observations together with errors induced by the model results. Here we may nevertheless benefit from well defined parameters used in the previous analysis: e.g. the bias and std. dev. of the geopotential altitude of the 100 hPa level as well as the height increment from 100 to 30 hPa and its std. deviation, at 00 and 12 UTC. This analysis gives some valuable information about the radiosondes' performances as they are operated by the different NMHSs (quality of operational upper air observations). Oakley compares the monitoring results for the main radiosonde types in section 3.1.3 of his 1998 report : "In general about 70% of the radiosonde stations in the period 1995 to 1997 were producing observations within the suggested quality limits, this compared with 65% in 1992". In the 2003 Elms report (Elms, 2003), in section 5.1.2, a worsening of the performance of many stations appears compared to the previous period, but its

conclusion in section 6.4 states that "the overall measurement quality of the radiosonde network has continued to improve".

(3). Time evolution of selected statistical parameters that would be recorded from the NMHSs in a world wide questionnaire.

A detailed questionnaire could be elaborated, in order to highlight the time evolution for radiosonde measurements worldwide. For example, statistical values such as the percent of soundings reaching the 100, 30 and 10 hPa levels, for the day and the respective night temperature and geopotential differences at 100, 30 and 10 hPa, could be helpful values in order to trace this improvement over the years. This would require an important contribution from the NMHSs. Height improvements could be a function of balloon improvements, size selection, or national needs.

(4). Compilation of the articles published in the scientific journals.

Information directly taken from the open literature may be another source of information about such global criteria for tracing the improvement of radiosonde performances over the years. A number of intercomparison campaigns and reviews have already been published for research projects such as ALPEX, TOGA, climate, and others. This fourth possibility could also represent an extension of the previous ones. Phase II report gives a list of radiosondes comparisons before the 1980s, the very first ones at Payerne in 1950 and 1956.

As a first approach we will use **the previous IOM reports in order to define some statistical criteria on measured data** (Option 1). The other options will be explored later if needed. Note that the next IOM reports on the compatibility of radiosonde geopotential measurements should allow a continuous analysis over more than 10 years. If the quarterly statistical parameters can be worked out for a longer period (e.g. 1988) or recalculated on the basis of ERA40, they would represent a valuable complement to the results obtained from the radiosonde intercomparisons.

3. Candidate quality performance criteria

Quality performance criteria should be based on well defined references. As absolute references are not or are only partially available for upper air measurements even in radiosonde comparisons, the "references" for calculating biases and standard deviations are based on the data obtained using as reference the "link radiosondes" or "high quality radiosondes" [2]. Candidate criteria are presented in Table 1. They always rely on comparison of simultaneous time-paired measurements. Comparison at standard pressure levels are not used here.

Table 1 shows a first selection of possible criteria: this selection is made with the objective of a small number of criteria in order to trace the improvement of radiosondes over the years with a straightforward data anaylsis. Criteria related to mean differences between sondes

[2] Since the "link radiosondes" were also in some cases showing intrinsic problems, another more general approach would be to use all the data measured under similar conditions in a given field campaign, and find the average and the standard deviation of the parameter under investigation. This has the advantage that no prior definition of a given "radiosonde link" is needed anymore. At this point of the study, this approach has not been followed..

(bias) include systematic measurement errors. They are more sensitive to radiosonde measurement problems than criteria based on the standard deviation of the differences between radiosondes. They allow determining the origins of radiosonde deficiencies. Standard deviation of the differences between radiosondes complements the information provided by the mean difference between radiosondes. If the standard deviation is smaller than the bias, it allows assigning measurement errors to a systematic problem in the radiosonde design and/or in the data processing. The standard deviation may also help raising error sources in the radiosonde reproducibility.

As a first example, the systematic temperature differences at 10 hPa in the night time are reported in Annexe I, as well as in Figure 1 below. The Excel sheet contains bias values extracted from Figures or Tables of the different IOM reports, without any additional processing. Each value is given with Figure or Table number as well as report number from where it has been taken out. Each symbol in Figure 1 corresponds exactly to a value reported in Table 1 (bias of one sonde type during one radiosonde comparison). Almost 30 radiosonde types have been engaged at least once over all comparisons. The radiosondes appear anonymously in this Figure. The aim is not to find out the best radiosonde, but to demonstrate that a general and continuous improvement in the quality of upper-air observations occurred over the last 20 years. The horizontal axis is a time axis covering the last 25 years. On the vertical axis, the span of the bias values is more important than their exact positions in relation to the zero point, as the reference is a relative one. In every comparison an outsider point may strongly increase the span. Basic statistical parameters may be added to the individual results. Nevertheless, they have a limited significance, as less than 10 radiosonde types (including different postprocessings for the same sonde) have been engaged in each comparison. Figure 1 contains a hand drawn envelope of the spans, where the first two comparisons are considered as a single one. In addition, the standard deviations of the biases are reported for each comparison, without combining the results of the two first comparisons.

Table 1. Candidate criteria for tracing the improvements of radiosondes

Criteria	**Remarks**
Temperature difference @10 or 30 hPa, @ night/day time	The 10 hPa level is the highest standard level in the TEMP messages. Meeting a high quality goal at this level is a demanding task. Temperature errors are different during night and daytime (noon). A higher data sample is found @ 30 hPa than @10 hPa.
Standard deviation of the temperature differences @ 10 or 30 hPa, @ night/day time	
Geopotential difference @ 10 or 30 hPa, @ night/day time	Geopotential measurements from radiosonde accumulate errors on other parameters (temperature, etc.) between surface and this level. Recent advances in GPS positioning have brought major upgrade on this criteria.
Standard deviation of the geopotential differences @ 10 or 30 hPa, @ night/day time	
Geopotential difference @100 hPa, @ night/day time	The 100 hPa level is the primary level used in the quality control of upper air data based on comparison with numerical model outputs.
Standard deviation of the geopotential differences @100 hPa, @ night/day time	
Systematic relative humidity difference in the temperature range between -35 and -45 degree Celsius (only tropospheric values)	Or two ranges: between -20 and -30C as well as between -40 and -50C. For modern humidity sensors temperature ranges below -50 C and/or in the stratosphere can help documenting their high performance. In order to account for the different humidity ranges, a further selection in 3 classes should be made: e.g. below 25 %, 25 % – 75 %, above 75 % RH.
Standard deviation of the humidity differences in the temperature range between -35 and -45 degree Celsius (only tropospheric values)	

Nevertheless, they have a limited significance, as less than 10 radiosonde types (including different postprocessings for the same sonde) have been engaged in each comparison. Figure 1 contains a hand drawn envelope of the spans, where the first two comparisons are considered as a single one. In addition, the standard deviations of the biases are reported for each comparison, without combining the results of the two first comparisons.

Indeed a first look at Figure 1 shows an improvement of the quality of the night time temperature measurements over the 20 last years. The hand drawn envelopes of the individual biases narrows with time, especially when considering the first two comparisons as a single one. The combination of Phase I and II night time temperatures was already presented in Figure 5.3 of the final report of Phases I and II. A span of 3.5 C was reported, much less than the combined value of 6 C as found according to the individual reports. An outlier was removed in the final report; in addition the reference values provided by the link radiosondes were improved. Hence, values reported in Figure 1 have to be interpreted with some caution. According to this span, the improvement of the night time temperatures at 10 hPa over the last 20 years lies between approximately 5 (span I-II based on individual values of report I and report II / span VI = 6 C / 1.2 C) and 3 (span I-II based on final report I-II / span VI = 3.5 C / 1.2 C).

A few basic statistical parameters can be calculated with the individual values of Figure 1. The standard deviation may be appropriate for the small samples of bias values. Its time series is plotted with vertical black bars in Figure 1 (1.7, 1.3, 1.3, 0.9, 0.5, and 0.6). According to this statistical parameter a decrease of a factor 2 to 3 might be attributed to the night time temperature bias at 10 hPa.

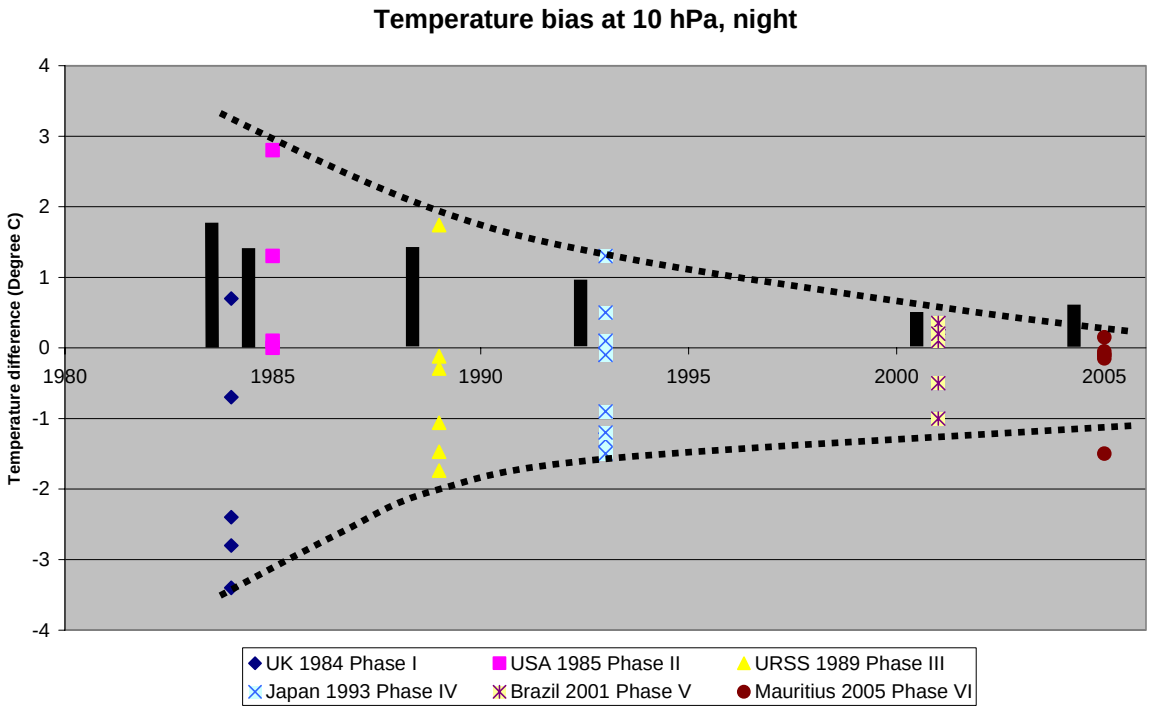

Figure 1. Night time temperature bias at 10 hPa for the 6 WMO Radiosonde Comparisons (simultaneous measurements). The two dotted lines represent a qualitative envelope of all individual results, considering the first two Comparisons as a single one. The vertical full bars correspond to one standard deviation of the biases of each comparison.

The analysis of this first candidate criterion brings valuable results, with the following comments:

- There is only "link radiosondes": as a consequence only relative numbers may be extracted from the comparison reports.
- The comparison reports for the six Phases present their results in rather different forms. For two campaigns, the final report is not yet available.
- Some comparisons were in particular addressing a certain class of parameters and thus do not present results for the other ones.
- There are not only different radiosondes, but different data post processing (correction of the radiation error on temperature, etc.).
- A single temperature criterion (e.g. 10 hPa bias for night time conditions) does not allow a comprehensive evaluation for temperature, although it refers to a very challenging altitude level. Other temperature criteria have to be addressed, e.g. for the low stratospheric and tropospheric temperatures.
- This type of analysis needs an additional method in order to provide technical and physical explanations, as well as a confirmation of its results. Therefore, in the following sections, some key interpretations and conclusions will be extracted from the comparison reports, and references to available results of process oriented studies will be given.

The following sections check some of the proposed criteria and present results for geopotential height, temperature, relative humidity and pressure.

4. Geopotential height

Up to a few years ago, radiosonde geopotential heights were mostly calculated with the hydrostatic equation. This method needs the pressure, temperature and humidity profiles and combines their errors. Nowadays, the newer radiosondes use the GPS technology, directly measures geometric height and converts it to geopotential height. This new technology brings a large improvement for the radiosonde pressure and geopotential height measurement accuracy.

Geopotential altitude measurements are highly demanding. We may remind that a 1 hPa error at 10 hPa corresponds to a 600 m geopotential error. Although reporting meteorological parameters at pressure levels (considered as true values) alleviates the errors of radiosonde measurements, this study is devoted to the accuracy of radiosonde measurements and only compares truly simultaneous measurements.

Figure 2 illustrates these improvements by pointing out in an anonymous manner all individual radiosonde bias of the geopotential altitude at the 10 hPa level, using the method described in section 3. In Phase II high precision radar was used as altitude reference. In Phase V, the first GPS radiosondes were introduced and the comparisons included in Figure 2 use GPS results as reference. Only GPS radiosondes of Phase VI are presented in Figure 2.

The hand drawn envelope on Figure 2 started with a span of 1000 – 2000 m in the first three comparisons. In Phases IV, better sensors and better calibration curves reduced this span to approximately 800 m. The move to the GPS technology brought a new standard in geopotential measurement accuracy. The GPS radiosondes are nowadays able to measure geopotential altitude at 31 km with differences of less than 20 m (cf. Mauritius Island experiment).

Phases I and II reports focused the main conclusions on geopotential altitudes up to 100 hPa. In addition, they gave some information on the higher layers, but fewer radiosondes provided measurements up to the 10 hPa level. Consequently, this hampered the statistical results at this highest altitude. The 2005 Maritius comparison is the first one that definitely demonstrates the accuracy jump brought by the GPS technology.

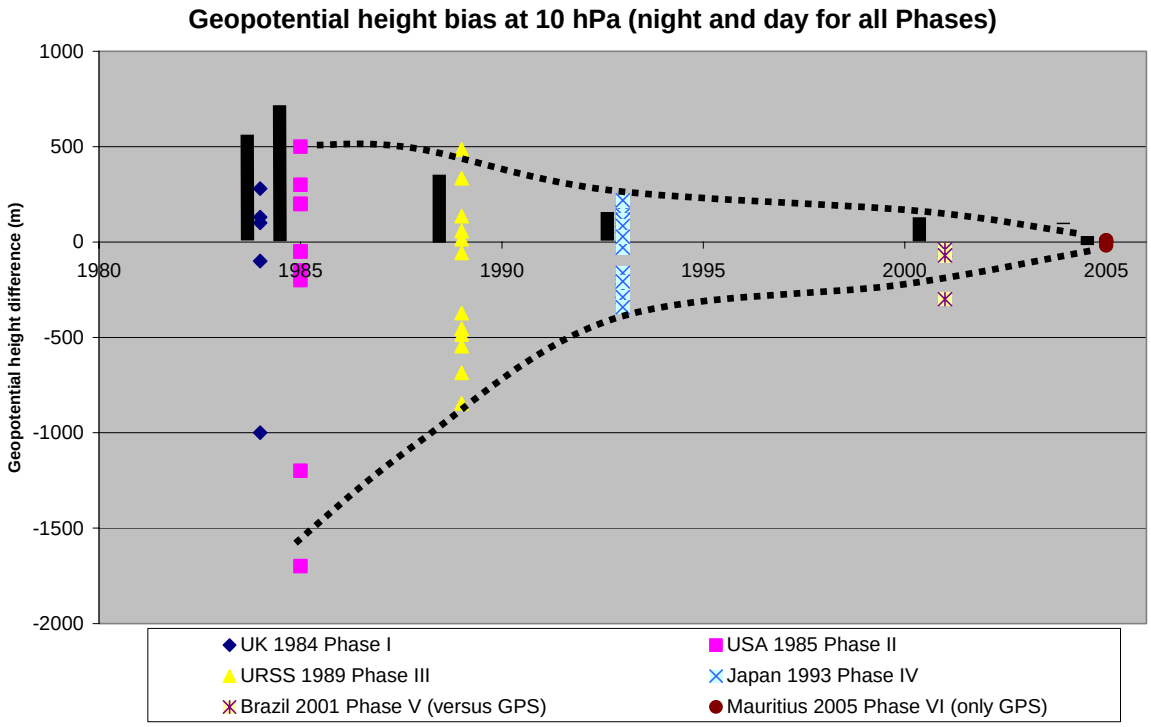

Figure 2. Bias of the geopotential altitude at 10 hPa for the 6 WMO Radiosonde Comparisons (simultaneous measurements). The two dotted lines represent a qualitative envelope of all individual results, considering the first two Comparisons as a single one. The vertical full bars correspond to one standard deviation of the biases of each comparison.

Within the last 20 years, the GPS technology jump has allowed improvements of an order of magnitude in the accuracy and reproducibility of radiosonde geopotential heights. This improvement has led to the additional conclusion in the IOM report [Nash, 2005]:

> "Thus, as long as the temperature sensor provides uninterrupted measurements of good quality, and the measurement locks correctly onto surface pressure and station height at launch, the cost of GPS radiosondes can be reduced by not using a pressure sensor."

Table 2. Results of the different WMO radiosonde comparisons related to geopotential heights.

Phase	**Results**
I, 1984	Three radiosondes agree within 30 gpm close to the ground, within 50 gpm at 16 km altitude, and within 230 gpm at 30 km. Two radiosondes run out from these data. Variability at 16 km ranges from a standard deviation of ±20 to ±140 m.
II, 1985	Agreement between radiosonde derived heights and radar measured heights (reference) was relatively good at levels below 100 hPa but differences became progressively larger above 100 hPa.
I + II	At 100 hPa the geopotential height observations of the radiosondes of 3 countries were on average about 50±20 m higher than the radar heights. At about 30 hPa the average bias between geopotential height observations of the radiosondes of two countries and the radar heights was always within the range ±100 m. At 32 hPa the day-night difference in the comparison of two radiosonde geopotential and radar heights was about 140±70 m, respectively 64±26 m.
III, 1989	The simultaneous geopotential height differences between three radiosondes do not exceed 30-40 gpm below 100 hPa and increase to 100-200 gpm at the 10 hPa level and night and up to 600 gpm (only one radiosonde) at daytime.
IV, 1993	In the nighttimes, differences among the radiosondes were within 150 m up to the 10 hPa level, except for two radiosondes. At midday, differences among the radiosondes were within 100 m up to the 10 hPa level, except for two radiosondes (due to differences in temperature or pressure).
V, 2001	Pairs of radiosondes showed height differences <100 m for the complete sounding. The other pairs showed an increase of the difference with the height, reaching up to 300 m.
VI, 2005	Once corrected for incorrect values of local g for the conversion from GPS heights to geopotential height, agreement between the GPS heights was excellent with estimated random errors lower than 10 m for most of the time and heights agreeing on average to ±20 m.

5. Temperature

In the 1980s temperature sensors were still of very different types: bimetal element with mechanical link, coiled tungsten wire (resistive element), thermistor rod, thermistor bead, thermocapacitive bead, etc. Later on the first two types were not used in the WMO comparisons and new sensors appeared.

Due to the different night- and day-time behaviour of the radiosonde temperature sensor because of the sunlight and infrared radiation, the radiosonde intercomparisons were systematically performed under both conditions (night and day). As far as possible, they also capture the daily cycle with soundings at different sun elevations.

Figure 3 illustrates these improvements by pointing out in an anonymous manner all individual radiosonde bias at the 10 hPa level. Both midnight and daytime results are presented on the same graph, in order to emphazise the long-term evolution of the envelope of

all negative and positive bias. The differences between Figures 1 and 3 are due to the inclusion of the daytime results.

Within the last 20 years, the technology evolution of the temperature sensors as well as more sophisticated data processing (calibration fits, removal of statistical bias, etc.) has allowed large improvements. Nevertheless, they are not as large as in the case of the geopotential measurements. An improvement of a factor between 2 and 3 emerges from Figure 3.

The following Table 3 traces the evolution of the temperature radiosonde performances along the different intercomparisons, by quoting key statements of their reports. Contrary to Figure 3, this Table encompasses tropospheric and stratospheric results. As in Table 2, it is worth mentioning that the first comparisons provided rather few results in the middle stratosphere.

Figure 4 [J. Nash, 2004] illustrates the influence of some changes or problems in temperature measurements, on the basis of the example of the Vaisala RS80 radiosonde. Similar process analyses represent the only way to explain how the improvements of radiosonde preformance have been reached. This is also the appropriate way to document and re-evaluate historical time series on a physical basis.

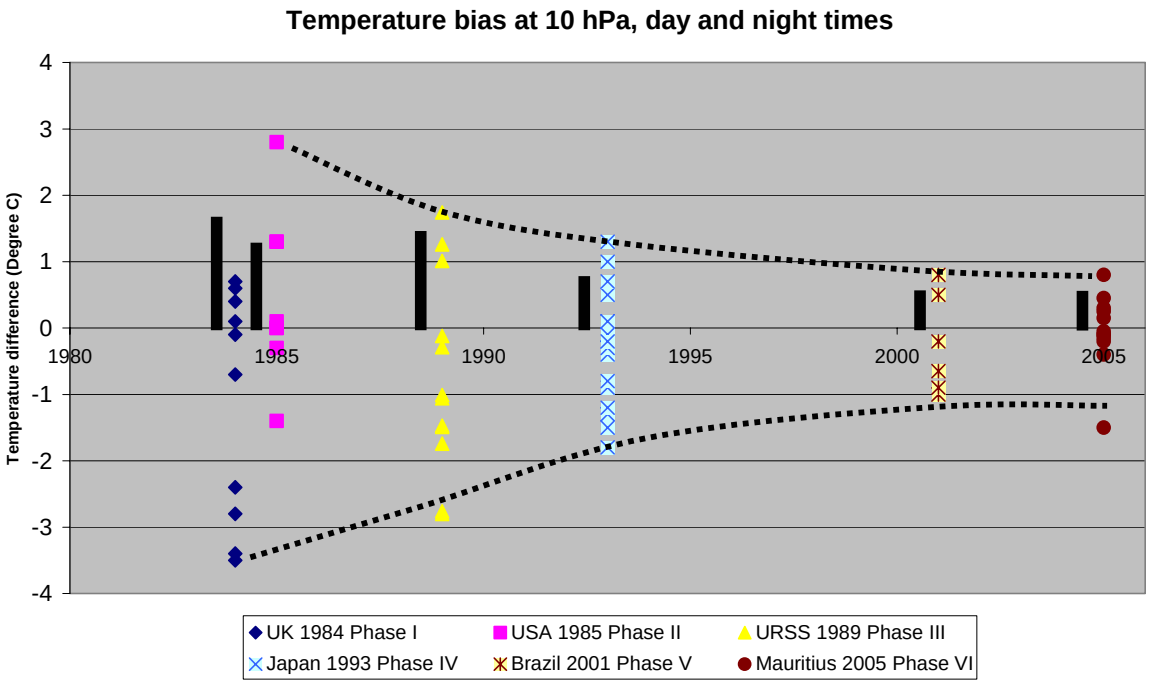

Figure 3. Night and day time temperature bias at 10 hPa for the 6 WMO Radiosonde Comparisons (simultaneous measurements). The two dotted lines represent a qualitative envelope of all individual results, considering the first two Comparisons as a single one. The vertical full bars correspond to one standard deviation of the biases of each comparison.

This analysis traces the continuous improvements in the quality of upper-air temperature measurements. The conclusions of Phases VI report pose other interesting questions, e.g. what are the performances (accuracy, reproducibility) of temperature sensors of the newest generation radiosondes? These conclusions are reproduced in Table 3 (see bottom lines). The conclusions previously given by J. Nash [2004] on "accuracy of best radiosonde temperature measurements" are reproduced here:

- Accuracy depends on temperature sensor error and also the error in the height (pressure) assigned to the temperature.
- Temperature sensor errors are smaller at night, as long as sensor coating has low emissivity in the infrared.
- Solar heating introduces significant systematic errors, difficult to correct, at pressure lower than 100 hPa.
- Random errors in temperature are less than 0.2 K at night and less than 0.3 K in daytime in the troposphere and lower stratosphere."

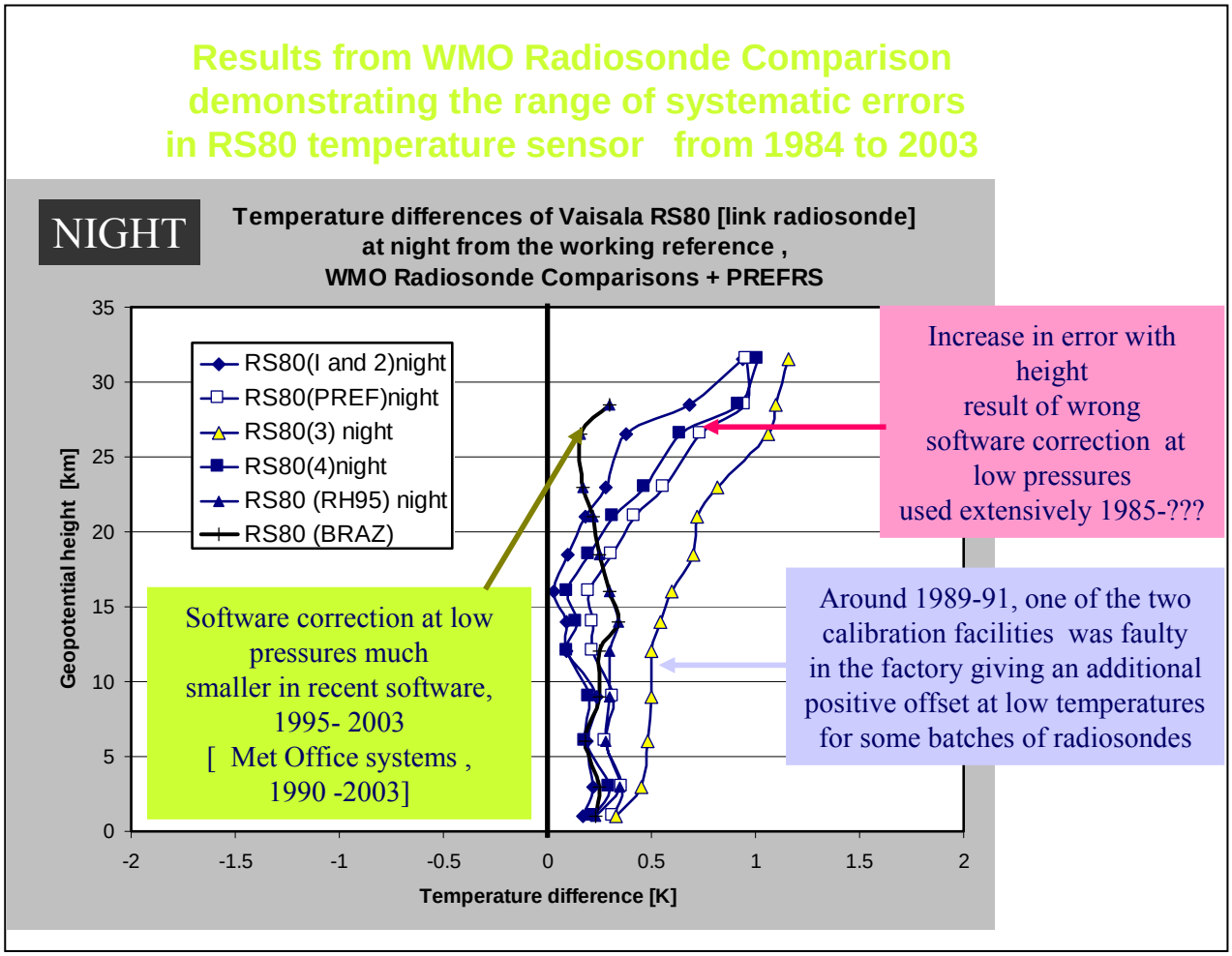

Figure 4. Range of systematic errors in RS80 temperature measurements from 1984 to 2003 [Nash, 2004].

Table 3. Results of the different WMO radiosonde comparisons related to temperature.

Phase	**Synthetic Results**
I, 1984	In the darkness the five radiosondes provide mean temperatures through the troposphere within 0.5 C of each other.
II, 1985	During the daytime the range of differences experienced by all of the sensors between surface and 100 hPa was less than 0.5 C. At night the range of differences increased to about 1.0 C.
I + II	The temperature at a given flight time could be measured to a reproducibility (1σ level) of at least 0.2 °C by most of the radiosondes. The observed temperature differences between radiosonde measurements were as large at night as in daytime conditions. At levels above 20 hPa the radiosondes with ducted temperature sensors show marked low biases at night, as also did the radiosondes using rod thermistor temperature sensors coated with lead carbonate. Certain discrepancies remain to be resolved (e.g., the range of 0.6 °C of the mean differences of the nighttime temperature measurements at levels below 50 hPa).
III, 1989	Observed differences between simultaneous temperature radiosonde measurements were within the range of 0.5 C, and in many cases 0.1-0.2 C, in the layers below 50 hPa. At levels above 30 hPa mean bias increased up to 1-2 C.
IV, 1993	In the nighttimes, temperatures measured simultaneously agreed within 0.3 C up to the 70 hPa level. Above the 70 hPa level, the differences among them increased to a maximum of 3.2 C, but temperatures obtained from aluminium-coated thermistor without infrared correction agreed well with those obtained from the 3-thermistor radiosondes within about 0.5 C. In the daytime, temperature obtained after the solar radiation correction agreed within 0.3 C up to the 70 hPa level, and their differences from those obtained from 3-thermistor radiosondes were within 0.5 C. Above the 70 hPa level, the differences among the temperatures obtained after the solar radiation correction increased and exceeded of 1.0 C, while the difference between the temperatures obtained from 3-thermistor radiosondes reached 0.7 C.
V, 2001	For daytime conditions, up to an altitude with a temperature of -30 C, the absolute differences among the radiosondes are within the range from -0.5 to +0.5 C. For temperatures lower than -35 C, the differences between radiosondes increased substancially. For nighttime periods, the differences are mostly reduced to a value close to -0.3 C up to the height of temperature -30 C. Above this height, the differences are still lower compared to daytime conditions and are mostly in the range from -0.5 to +0.5 C:
VI, 2005	Random errors in daytime temperature measurements of the two best radiosondes were less 0.2 C at heights up to 30 km, whereas random errors in the other temperature sensors were larger than 0.2 C at heights above 16 km. Two systems offer best performance night and day over the complete altitude range (most suitable radiosondes for climate monitoring up to 5 hPa). All participating systems are acceptable for good quality operational use for flights up to 30 hPa. It is doubtful whether “reference” quality radiosondes can provide more consistent measurements than the best operational radiosondes.

6. Relative humidity

In the early 1980s, goldbeater's skin, hair and Lithium Chloride sensors were still in use in some radiosondes, while others had already moved to carbon film or organic polymer film hygristors. In the mid 1980s carbon resistive hygristors were in widespread use with good results in the low troposphere. Further developments on thin film capacitor sensors have brought them to the most suitable operational humidity sensor for the lower and upper troposphere. Nowadays this technique is used in all modern operational radiosondes. The chilled mirror hygrometer remains a reference sensor for research and development studies

The humidity measurements underwent large improvements in the last decades, but accuracy and reproducibility goals are still challenging goals for the low temperatures encountered in the upper troposphere. The difficulty of upper air humidity measurements is well illustrated by the fact that they have been limited for many years to the low and middle troposphere, for temperatures above and slightly below zero Celsius. Phase I and II radiosonde intercomparisons restricted the humidity performance analysis to the 500 or 400 hPa levels. Later on, some of the intercomparisons were especially dedicated to humidity and additional intercomparisons were organized for this parameter, e. g. at Wallops Flight Facility in Sept. 1995. Phase V intercomparison in Brazil and Phase VI in Mauritius Island were focusing on humidity sensors.

It is known that accurate measurements of relative humidity are influenced by pressure, temperature, solar radiation and contamination from water drops on the sensor. Moreover, the relative humidity can change rapidly, especially when the radiosonde passes through a cloud. The relative humidity is a measure that is referenced to saturation (100 % RH). As the latter strongly depends on temperature, a 50 % RH value at 20 C corresponds to much higher water vapour content than at -50 C. This wide range of water vapour content in the atmosphere implies that the humidity performance analyses should also include a parameter such as the specific humidity or the mixing ratio.

The WMO radiosonde comparison reports do not present humidity results in a uniform way that would allow the extraction of comparable results for all Phases. Systematic relative humidity difference in the temperature range between -35 and -45 degree Celsius (only tropospheric values) can be found in only some of the reports. The classification changed with time in these reports. Pressure or altitude classes were used in the first ones, which are not appropriate for humidity (saturation depends on temperature, not on pressure). Alternatively, the analysis was restricted to one class with temperatures between 0 and +20 C. The classification has been improved along the different reports. Nevertheless, one should reanalyze the datasets and recalculate statistics according to a single scheme.

Alternatively, the synthetic results can be extracted from the WMO comparison report (Table 4).

Table 4. Results of the different WMO radiosonde comparisons related to relative humidity.

Phase	**Synthetic Result**
I, 1984	At high relative humidity the standard deviation of all five designs is typically 2.5 % RH, increasing to 6 % RH for humidity below about 50 % RH.
II, 1985	Based on a repeatability of 2 % RH for the carbon hygristor, the repeatability appears to be 4-6 % for the capacitive sensor, and 10 % for the LiCl hygristor.
I + II	The carbon hygristor sensor has a typical reproducibility of about 3.5 % RH, but a poor resolution at RH below 20 %. The thin film capacitance sensor measures too low near saturation in low level clouds, but is considered more reliable than the carbon hygristors at the dry end of the humidity scale. Goldbeater's skin, hair and Lithium Chloride sensors have more limited capabilities than the carbon resistor and thin film capacitor sensors.
III, 1989	The thin film capacitor sensor had a better time response at lower pressure than the other sensors. However, it did not prove the same reliability under pressure significantly lower than 200 hPa.
IV, 1993	Large humidity differences were observed in the low humidity range, according to the type of sensor (capacitive film or carbon hygristor).
1995	None of the sensors reported identical humidity profiles. A final, and very important conclusion is that it is doubtful that the sensor measurements can be accepted at temperatures lower then -40 C.
V, 2001	In the troposphere up to around 8000 m, where the mixing ratio is large, the radiosonde measurements presented a low dispersion. At higher altitudes the measurements were highly dispersed.
VI, 2005	*Waiting for final report*

Completion of this section requires the Mauritius Islands report, as well as systematic calculations on the basis of the original datasets.

7. Pressure

The pressure sensors in the early 1980s were based on the measure of the deflection of the plates of an aneroid cell. However, the methods applied for transferring pressure change with respect to altitude change were quite different. Hypsometers were mainly used with ozonesondes. Next generation pressure sensors are e. g. silicon chips, water hypsometers.

To be completed

Table 5. Results of the different WMO radiosonde comparisons related to pressure.

Phase	**Synthetic Results**
I, 1984	Three radiosonde designs provided pressure data within, on average, one hPa of each other in the troposphere, reducing to 0.1 hPa of each other at the 10 hPa level. Two other designs differed substantially from that of the first three designs. The pressure variability (standard deviation) of the best two designs was less than ± 0.75 hPa in the lower troposphere, reducing down to within ± 0.35 hPa in the stratosphere.
II, 1985	Above 500 hPa, comparisons show the mean pressure difference to be less than 0.5 hPa.
I + II	Significant (greater than 2 hPa) mean bias errors existed in the pressure measurements of three of the eight participating radiosondes.
III, 1989	The range of values indicated by different systems was 2-3 hPa at 500 hPa decreasing to 0.3-0.5 hPa at high levels.
IV, 1993	The biases of most radiosondes were within 1.5 hPa up to the 200 hPa level.
V, 2001	In general, differences between pairs of radiosondes are in the range of -2.7 up to +2.7 hPa in the first layer (surface – 8 km), reducing this to -0.9 up to 1.1 hPa in the third layer (above 8 km).
VI, 2005	*Waiting for final report*

8. Note to the CIMO upper air expert group

This report represents a first answer to the action initiated at the 2004 CIMO expert team meeting on upper-air systems intercomparisons (ET on UASI-1, Geneva, 17-20.03.2004): "Elaborate global criteria for tracing the improvements, based on previous intercomparisons and recent radiosonde development, and including remote sensing". It identifies four possible approaches in order to define these criteria, and selects option 1 as a first result for the 2005 expert team session (28-30.11.2005):

(1). Time evolution of selected statistical parameters reported in the different IOM international radiosonde comparisons.

(2). Time evolution of selected statistical parameters reported in the different IOM analyses on the compatibility of radiosonde geopotential measurements (comparisons with numerical weather prediction first-guess).

(3). Time evolution of selected statistical parameters that would be asked from the NMHSs in a world wide questionnaire.

(4). Compilation of the articles published in the scientific journals.

A list of candidate performance criteria is proposed as well as the way to extract the corresponding values from the IOM reports and to present the results. Some comments on the possibilities and limits of this method are already included in this first report.

Geopotential height is the first selected criterion with clear improvements over the two last decades, as shown in Figure 2. The GPS technology allowed an improvement of an order of magnitude in the quality of radiosonde geopotential heights at 30 km altitude. Table 2 gives some additional information on the validity of the results presented in Figure 2. As a consequence of the results of the Mauritius Island experiment, clear statements regarding the future of radiosonde pressure measurements were stated.

Large improvements have been achieved for **temperature** as shown in Figure 3: an improvement of a factor of 2 – 3 at 30 km altitude is reported. Table 3 gives some indepth information on the validity of Figure 2. The detailed study shown in Figure 4 is a complementary study. Not only the comparison reports would benefit from a more uniform data analysis and presentation, it is also of primary importance that the related database remain accessible worldwide. It is worth noting that a lot of effort has already been devoted to these goals.

Upper-air **humidity** measurement is a challenge. Humidity has been only partially treated in this report. Completion of this section needs the final reports of the last two radiosonde comparisons and new systematic calculations on the basis of the original datasets.

The **pressure** section of this report is briefly mentioned and **wind** is still missing in this report.

Final comments:

This report focuses on the past 20 years or so. Its main objective is to demonstrate in a simple approach the radioosnde achievements over the last 20 years. Before completing this report, a feed back is asked from the expert group:

- Review of the objectives of this task.
- Proposed criteria: guidance for completion.
- Decision on the type of publication

Next steps after the review of the expert group (November 2005):

- Complete report in spring 2006, according to the editor's requirements.
- Last reviewing by the expert group.
- Publication

9. References

….

10. Annex

Table. Example of Excel sheet used for the extraction of the quality performance results of the Phases I to VI WMO radiosonde comparisons. The source of each result is referenced with a figure number. In addition, the reference used in the intercomparisons, the report and figure are given. In the first reports, the graphs do not present the corresponding values: a direct "analogue" reading of the values on the plots is then specified (in which case the absolute value reported may already be affected by this step!).

Systematic temperature difference at 10 hPa, night time in Degree Celsius	UK 1984 Phase I		USA 1985 Phase II		URSS 1989 Phase III		Japan 1993 Phase IV		Brazil 2001 Phase V		Mauritius 2005 Phase VI	
Sonde	1984		1985		1989		1993		2001		2005	
OCAN 1524-511	**-2.8**	4a										
RS 3 (UK)	**-0.7**	4a										
RS4 MK3(Aus)			**0.1**	5.8								
MK-III (India)			**1.3**	5.8								
Graw 78 C (D)	**-3.4**	4a										
Graw DFM97 (D)									**0.35**	ppt	**-0.05**	3.a
SMA-TC-1 (SMT)												
SMA-GZZ (SMG)					**-1.47**	5.9						
MARS-2 (MRS)					**-0.12**	5.9						
MRZ-3A (MRZ)					**-0.29**	5.9						
Meisei RS2-80 (JP1)							**-0.9**	2.1a				
Meisei RS2-91 (JP2)							**0.1**	2.1a				
Meisei RS-01G							**0.1**	2.1a			**-0.1**	3.a
Vaisala RS80-15N (FN1)	**0.7**	4a	**2.8**	5.8	**1.74**	5.9	**1.3**	2.1a				
Vaisala RS80-15N (FN3)							**0.5**	2.1.c				
Vaisala RS80-15LH (FN2)									**0.35**	ppt		
Vaisala RS90-…									**0.1**	ppt		
Vaisala RS92-…											**0.15**	3.a
AIR IS-4A- (AR1)					**-1.06**	5.9	**-1.3**	2.1a				
AIR IS-4A- (AR2)							**-1.3**	2.1a				
AIR IS-4A- (AR3)							**-0.1**	2.1.c				
VIZ 1392 (VIZ0)	**-2.4**	4a	**0**	5.8	**-1.74**	5.9						
VIZ Mark II (VIZ)							**-1.2**	2.1a		6.3.1.b		
VIZ Mark II (VZ2)							**-1.5**	2.1.c				
VIZ Mark II (VZ3)							**-0.1**	2.1.c	**-0.5**			
Sippican LMS-5									**0.2**		**-0.1**	3.a
GL-98 (MODEM)									**-1**	6.3.1.b	**-1.5**	3.a
SRS-C34											**-0.15**	3.a
Reference used:	Mean(FIN,UK)		VIZ0		Mean(FN1,VIZ0)		Mean(FN1,AR1)		id. Mauritius		Mean(Me, Si, Va)	
Table used:	I:4a		II: 5.8 (2300UTC)		III: 5.9		IV: 2.1a, 2.1.c		ppt Nash		Fig. 3a	
Type of reading	Analogic		Analogic		Digital		Digital		Analogic		Analogic	

Same as previous Table, but for the daytime and with some statistical parameters.

Systematic temperature difference at 10 hPa, day time in Degree Celsius	UK 1984 Phase I		USA 1985 Phase II		URSS 1989 Phase III		Japan 1993 Phase IV		Brazil 2001 Phase V		Mauritius 2005 Phase VI	
	1984		**1985**		**1989**		**1993**		**2001**		**2005**	
Radiosonde												
OCAN 1524-511	**-3.5**	**4d**										
RS 3 (UK)	**-0.1**	**4d**										
RS4 MK3(Aus)			**-0.3**	**5.5**								
MK-III (India)			**1.3**	**5.5**								
Graw 78 C (D)	**0.4**	**4d**										
Graw DFM97 (D)									**-0.65**	**ppt**	**0.45**	**4.a**
SMA-TC-1 (SMT) China												
SMA-GZZ (SMG) China					**-2.76**	**5.10**						
MARS-2 (MRS) URSS					**-1.49**	**5.10**						
MRZ-3A (MRZ) URSS					**-2.8**	**5.10**						
Meisei RS2-80 (JP1)							**-1.8**	**2.2a**				
Meisei RS2-91 (JP2)							**0.1**	**2.2a**				
Meisei RS2-01G											**0.8**	**4.a**
Vaisala RS80-15N (FN1)	**0.1**	**4d**	**-1.4**	**5.5**	**-1.01**	**5.10**	**-0.8**	**2.2a**				
Vaisala RS80-15N (FN3)								**2.2.c**				
Vaisala RS80-15LH (FN2)									**-1**	**ppt**		
Vaisala RS90-…									**-0.2**	**ppt**		
Vaisala RS92-…											**-0.2**	**4.a**
AIR IS-4A- (AR1)					**1.26**	**5.10**	**0.7**	**2.2a**				
AIR IS-4A- (AR2)							**-0.4**	**2.2a**				
AIR IS-4A- (AR3)							**-0.3**	**2.2.c**				
VIZ 1392 (VIZ0)	**0.6**	**4d**	**0**	**5.5**	**1.01**	**5.10**						
VIZ Mark II (VIZ)							**1**	**2.2a**				
VIZ Mark II (VZ2)							**-0.1**	**2.2.c**				
VIZ Mark II (VZ3)							**-0.2**	**2.2.c**	**0.8**	**ppt**		
Sippican LMS-5									**0.5**	**ppt**	**0.25**	**4.a**
GL-98 (MODEM)									**-0.9**	**ppt**	**0.3**	**4.a**
SRS-C34											**-0.4**	**4.a**
Statistics												
Minimum	**-3.5**		**-1.4**		**-2.8**		**-1.8**		**-1.0**		**-0.4**	
Maximum	**0.6**		**1.3**		**1.3**		**1.0**		**0.8**		**0.8**	
Span	**4.1**		**2.7**		**4.1**		**2.8**		**1.8**		**1.2**	
Mean	**-0.5**		**-0.1**		**-1.0**		**-0.2**		**-0.2**		**0.2**	
Mean Abs. Dev.	**1.2**		**0.8**		**1.4**		**0.6**		**0.6**		**0.3**	
Sigma	**1.7**		**1.1**		**1.8**		**0.8**		**0.7**		**0.4**	
Mean - 1*sigma	**-2.2**		**-1.2**		**-2.7**		**-1.0**		**-1.0**		**-0.2**	
Mean + 1 sigma	**1.2**		**1.0**		**0.8**		**0.6**		**0.5**		**0.6**	

Reference used:	Mean(FIN,UK)	VIZ0	Mean(FN1,VIZ0)	Mean(FN1,AR1)	id. Mauritius	Mean(Me, Si, Va)
Table used:	I:4d	II: 5.5 (1400UTC)	III: 5.10	IV: 2.2a, 2.2.c	ppt Nash	Fig. 4.a
Type of reading	Analogic also p84	Analogic	Digital	Digital	Analogic	Analogic

Sigma daytime-nighttime	**1.7**		**1.3**		**1.5**		**0.8**		**0.6**		**0.6**