12.5 **REAL TIME DATA MONITORING AT NCEP**

V. Krishna Kumar*, Bradley Ballish and Jeffrey Stoudt
NCEP Central Operations, Camp Springs, MD

1. INTRODUCTION

NCEP operations receive meteorological and oceanographic data from many data suppliers, 24 hours a day, in real time. Real time monitoring of data is therefore indispensable for ensuring the quantity and quality of these multi-source, multi-platform data, which are critical for ingesting into all the operational NCEP models for initialization. Considering this critical role, NCEP developed a web-based Real Time Data Monitoring System (RTDMS) to ensure both normal quantity and quality of the data. Since data dump counts are already produced operationally for model runs and for hourly time periods, Phase I of this system would principally focus on data counts and would be the first phase to become operational. The next phase, Phase II would then focus on the quality of data, other than satellite radiance data, as well as upgrades to Phase I. Phase III would focus on satellite radiance data using diagnostics from the Joint Center for Satellite Data Assimilation.

Phase I of NCEP's Real Time Data Monitoring System (RTDMS) became operational in the NCEP production suite in July 2005. This data monitoring system compares current and monthly average counts of many different observational data types for different model runs as well as for hourly time periods. This web-based system consists of various color summary pages and detailed time series graphics of data counts of all data types. This system enables NCEP 24x7 operational staff to quickly discover what data counts are abnormally low or high so that they could take remedial action or notify data suppliers of a specific problem in a suite of operational models. The URL for this site is:
http://www.nco.ncep.noaa.gov/pmb/nwprod/realtime/

* Corresponding author address:
Dr. V. Krishna Kumar, NCEP Central Operations, Room 302, World Weather BLDG., 5200 Auth Road, Camp Springs, MD 20746. e-mail: Krishna.Kumar@noaa.gov

Phase I of the RTDMS focuses primarily on the quantity of all data types that NCEP receives. Efforts are currently underway at NCEP for developing technique for the Phase II of the RTDMS, which will highlight not only the quantity of the data but the quality of all data types.

This paper is structured as follows. Section 2 and 3 describe, respectively, the specific technical details of the recently operational RTDMS Phase I and not yet operational RTDMS Phase II. Some of the challenging issues in developing comprehensive time series Quality Control (QC) procedures for Phase II along with some preliminary results are discussed in Section 3. Section 4 outlines a brief summary with future plans of this evolving system.

2. DESCRIPTION OF THE OPERATIONAL RTDMS – PHASE I

Phase I compares current data counts with their monthly average counts of many observational data types that were available for NCEP model runs and for around-the-clock hourly time periods. For data suppliers, it may be extremely valuable to monitor this site to ensure that their specific data arrived at NCEP as expected in a timely fashion. This is especially important if the supplier has made a system change or knows of any abnormality. It is possible for the data supplier to detect problems before NCEP.

More specifically, the main page of this web-based system provides the options to select data availability by clock hour (UTC or Z time) or by different NCEP model run times during the last 24 hours. The monitored NCEP models (three letter acronyms used) include the North American Mesoscale (NAM), Global Forecast System (GFS), GFS Global Data assimilation System (GDS), and the hourly Rapid Update Cycle (RUC) runs. The boxes that correspond to the most recently available hour and model cycles are shaded which imply the data processing and graphics have been completed and available for the users. After users select the desired option, they are presented with a color-coded summary table that shows the monitored data types by mnemonic, see Fig. 1 for example. Users may reference 'Explanation of Data Types' for a brief description of any mnemonic.

This page also provides an explanation of the color codes. Based on the importance and the reliability of various data types, some data types are considered critical or sub-critical for operational use in the NCEP models. For critical types, green indicates a count within normal range, red a shortage, and purple means an overage. For sub-critical types, green indicates a normal count, orange represents a shortage and blue indicates an overage. For sub-critical types in hourly counts, if the count is considered low for two hours in a row, then red is displayed for that data type. For some types of data whose counts are not as steady, they are labeled as "data of opportunity" with a display in black. For individuals who have difficulty viewing the colors, they may choose to view the "text summary" which is in color, but the color is not critical to read the messages. The "text summary" can also be useful as it provides text counts that may be more precise than what can be estimated from a time series plot, and, in addition, it provides time windows for the data dumps. For critical data types, if the current count is less than a specified fraction of the monthly count, then the count for that data type is considered to be low. The specified fraction varies with data types, but most data types, a value of 0.5 (50%) has been chosen.

Therefore, users should be able to quickly discern any abnormal shortages or overages in counts of critical and sub-critical data types. Then users can view time series plots such as Fig. 2 after selecting the desired data type mnemonic. These plots should show the latest trend of counts as compared to the monthly average, which may reveal the severity of any abnormality and when it began or returned to normalcy. For hourly counts, the counts show the last 24 hours, while for model runs, the counts show 5 to 7 days worth of counts depending on the model.

Monitoring of hourly data and 6-hourly dump (GFS, GDS) and 3-hourly dump (NAM) for models have distinct advantages. Users can opt to view data counts by model, especially when they are most interested in data availability for the selected model run. The time series graphs visually augment the printed text reports that become available to NCEP operational staff immediately prior to the execution of NAM, GFS, and GDS models. However, availability of per-model data counts is rather infrequent as compared to hourly as a few hours elapses between some of the adjacent model runs. Therefore, monitoring the hourly counts may result in detecting and resolving data count issues in time for the next NCEP model run.

Monitoring data counts and troubleshooting data related problems are job responsibilities of NCEP operational staff. NCEP staff has been trained to deal with data related issues. However, the 'Data Troubleshooting Guide' has been made available to reference for guidance in resolving data problems. The guide contains information on the source or suppliers of data and the network of data communication to the NCEP supercomputers as well as procedural tips including WMO bulletins or AWIPS products to check and links to suppliers or servers.

The following example is taken from a case dated 25 October, 2005. The hourly product at 05Z (see Fig. 1) displays a color red for dbuoy (drifting buoy), mbuoy (moored buoy), and lcman (land-based CMAN station) data types. The 24 hour time series ending 12Z on 25 October, 2005 (see Fig. 2) indicates the problem began at approximately 04Z where the depicted mbuoy and lcman counts fall sharply below the normal counts. There had already been an ongoing issue concerning dbuoy data. Additionally, the tideg (tide gauge) data, a "data type of opportunity", also show similar problems. According to the NCEP Senior Duty Meteorologist (SDM) log at 0420z 25 October 2005, the SDM noted that "Wallops communications are having some issues," and that Ocean Prediction Center (OPC) forecasters were reporting a "gradual decrease in buoy reports during the past hour." Later, at 06Z on 25 October 2005, the SDM noted, "Wallops found a bad power supply and then switched/fixed this problem." Later NCEP and NWSTG operational staff notified the data supplier at Wallops Island, VA, who quickly diagnosed and switched away from a faulty power supply. Indeed, the dbuoy, mbuoy, and lcman time series graphs each showed a recovery of data counts to near normal between 05Z and 06Z, which was in time for the 00Z GDS model that began at 0604Z.

3. TECHNICAL DETAILS OF THE RTDMS – PHASE II PLAN

For Phase II of the RTDMS, the key objective is to be able to monitor the quality of different classes of data used in model analyses. For example, if the quality of the AMDAR temperatures degrades suddenly, the Phase II web based system could alert the center's 24x7 operational staff that this data is suspect, displaying a color orange for this data type. If the data quality or quantity is judged to be a sufficient problem, the web page would show red and the 24x7 staff would be required to take appropriate action. For measuring quality, statistics on different classes of data compared to the NCEP guess are used. The basic idea is to compare current counts, biases and RMS Differences to the Guess (RMSDG) for different classes of data compared to their historical statistics. After several experimentations, it was decided that a 30-day running history would be a good choice for the time mean. Now if the current statistics of a particular data type from a particular model run differ by more than a pre-determined number of standard deviations (STD) from the running 30 day mean, the data would be considered suspect. Recall that Phase I of the RDTMS compared current dump counts to their monthly means and only used tests on the ratio of counts to decide if the count was suspect. In contrast, for Phase II, various rigorous statistical checks are required in part for developing a robust system. Our earlier investigation has shown that defining problematic classes of data by statistical checks alone results in too many false alarms. Therefore, some improvements are needed beyond the statistical checks.

It would be impractical and problematic to generate statistics for all data types at too many pressure levels in a real time operational environment. Based on our investigation, it was decided to restrict computing statistics to 5 distinct pressure categories. The first level is "surface" for data types valid at the surface. The next level is "low" for all data from the surface to 700 hPa, excluding the true surface data. The category "mid" is for data above 700 hPa and below 300 hPa. The category "jet" is for 300 to 200 hPa. Finally, "high" is for data above 200 hPa. This methodology could change in the future, but it allows sampling of different pressure categories without the number of categories getting too high or the quantity of data in each category being too low.

The definitions of all classes of data within the NCEP model Binary Universal Form for the Representation of meteorological (BUFR) data files are provided by D. Keyser and available at: http://www.emc.ncep.noaa.gov/mmb/data_processing/prepbufr.doc/table_2.htm.
For example, radiosonde temperatures and winds are assigned with numbers 120 and 220 respectively. ACARS temperatures are assigned with a number 133. Mass fields such as temperature are assigned numbers between 100 and 199, and momentum fields such as wind are always between 200 and 299. For tests described later for the month of March 2005, about 122 different categories of data were used, which result in 366 time series to check. For example, a specific data type such as radiosondes, has 16 categories due to temperature, moisture, and winds at 5 pressure levels along with surface pressure. For Phase II, the counts, biases and RMSDG for these 16 categories result in 48 time series checks, where as in Phase I, only one total count is checked. With roughly 366 time series QC checks to be performed at least 4 times per day, it is not surprising that earlier work resulted in too many false alarms. All observations or variables that had available NCEP guess values were used for computing statistics.

A number of experiments with the Phase II time series QC codes were performed for the month of March 2005. The data corresponding to March 2005 were particularly chosen as they had real problems with quality as well as quantity. Therefore, the QC codes could be fully tested in the presence of difficult problems and may be considered as a representative benchmark data set. Before comparing any of these tests, a fundamental change was required for time series QC checks on data counts. In March 2005, there was an abrupt increase in surface temperature counts for the data type 183. Figure 3 shows the daily counts for 00Z runs. The average count had been very steady at about 29 observations, and then on 3 March 2005, the count jumped to around 500 observations. This was a very large change in counts measured in STDs. Since it was not desirable to have the code reject these large counts thereby producing alerts for the center's 24x7 operational staff for a prolonged time, the code was modified so that if a count was rejected for being high or low for 3 runs in a row, the QC flags were adjusted to be suspect and the data allowed to be used in the 30 day running mean. This QC adjustment allows the code to pass counts with large consistent increases or decreases in about a few model runs with no human intervention and is used in forth coming tests.

A sequence of four statistical tests on the time series QC was conducted for the month of March 2005. Test1 treated any current count, bias or RMSDG to be considered Suspect Data (SD) if it differed by 3 STDs or more from the running mean. For differences of 4 STDs or more the data was considered Very Suspect Data (VSD). Such VSD would not be used in the next running mean as it could considerably alter the statistics. Test2 was the same as Test1 except the limits of 3 and 4 STDs were replaced respectively by 4 and 5. In Table 1 is shown the number of times that a data count is considered to be a Suspect Count (SC) or a Very Suspect Count (VSC), a Suspect Bias (SB), a Very Suspect Bias (VSB), a Suspect RMSDG (SR) and a Very Suspect RMSDG (VSR). For Test1, 210 cases of VSD were found. This is excessive and would be considered "crying wolf" too often. Test2 is doing better in terms of reducing false alarms, but is still considered excessive.

Analysis showed that one problem with the time series QC checks was that data categories with low current counts caused too many false alarms. Since a category with a low count has a relatively small effect on the analysis and since the low count makes the statistics less reliable, a Low Count Factor (LCF) was defined and used to reduce false alarms. This variable is defined as LCF = (1.0 + SQRT (100.0/ (COUNT + 1.0))), where COUNT is the current count of the data category. For example, if COUNT=24, then LCF=3.0. For counts on the order of 1000 or more, LCF is roughly one. A statistical check such as the mean +/- N*STD has factor N replaced by N*LCF. Test3 is the same as Test2, except the LCF was used in the statistical checks. As seen in Table 1, this greatly reduces the number of times that the data are flagged as SD or VSD. For an example of the need for using the LCF, Fig. 4 and 5 show, respectively, the counts and biases of low-level temperatures for AIREPS. Note, that if the count exceeds 5, that would be more than 3 STD away from the mean. This certainly does not require an alert to the 24x7 staff. On 23 March 2005, the bias of the above class of data is about 7 degrees, which is well beyond 3 STDs from the running mean (Fig. 5). This bias was due to just one observation and an alert to our 24x7 staff is unwarranted for this case. For Phase II, only major data quality problems should be alerted without attempting to check all the minor details pertaining to data QC.

In Test3, many data counts are considered suspect even though there was not a relatively large change in counts. What frequently happens with data counts is that the count in the running mean can be very steady with a small STD, making small changes in count look suspect. Figure 6 shows current counts of marine surface pressure data at 00Z from early February to mid-March, 2005, labeled as "CUR", while the running mean is labeled as "AVE". The statistical limits given by the running mean +/- 5 STDs are labeled as "MAX" and "MIN". On 4 March 2005, the count increases from roughly 4 to 5 thousand. This increase is about 6 STD over the mean, but is a relatively small percentage change. Thus, in Test4, a count was not considered suspect unless the current count minus the running mean was at least 50% of the running mean. This "ratio test" for the counts greatly reduced VSC from 54 in Test3 to 5 in Test4 (see Table 1). The final tuning options in Test4 resulted in only 12 alerts for all of March 2005.

Test4 results in 7 cases of data quality being flagged as VSD. For these cases, a sequence of forth-coming graphics show time series of the current stats "CUR", the statistical QC limits "MAX" and "MIN" as well as the running mean "AVE". Here, MAX and MIN are given by the running mean +/- 5*LCF*STD. The first case, Fig. 7 displays RMSDG for NESDIS satellite infrared cloud drift winds above 200 hPa for 00Z. For 10 March 2005, the current RMS wind difference goes well beyond the maximum limit. Analysis of the suspect data in this case shows a group of winds in the Pacific area around 110 hPa that has large differences from the guess. These winds agree very well with the same type of satellite winds in the same area but at 300 hPa. This appears typical of an error in the assigned pressure of the winds around 110 hPa. These suspect winds passed data QC and had some negative impact on the analysis. Figures 8 and 9 show respectively time series of RMSDG and speed bias for satellite winds with an isolated problem of wrongly assigned pressure levels. Figure 10 shows a problem with the root mean square wind differences for MODIS infrared satellite winds of type 257 for 06Z on 10 March 2005, which seems to be due to excessive number of calm winds.

Figure 11 shows that the marine winds were suspect at 18Z on 8 March 2005, while Fig. 12 shows that the marine surface pressure was suspect six hours later for 00Z 9 March 2005. These two cases appear to have suspect data due to the surface pressure changing very rapidly along the US east coast coupled with the model BUFR files not having time interpolation of the guess.

Thus, the code used in Test4 is able to detect some serious problems with data quality without making too many false alarms. Further testing and tuning of the code are likely before it is made operational. One possibility is to perform time series QC checks on consecutive model runs. For example, if the speed bias of a certain wind type is roughly 3 STDs from the running mean for 3 adjacent model runs, there may be a real problem with the data that does not meet the suspect criterion of Test4. Another possible improvement is to adjust time series QC decisions based on the criticality of the data class.

4. SUMMARY AND FUTURE PLANS

Phase I of the operational RTDMS is summarized and found to be useful for data providers and the NCEP staff. It is particularly useful for the 24x7 operational staff to detect problems in data flow or supply and correct them in a timely fashion prior to running the next operational models. Earlier time series QC statistical tests for the Phase II of the RTDMS resulted in too many false alarms. By adding a modification for data with low counts and a ratio test for data counts, the system was able to detect real problems with various classes of data resulting in a large reduction in false alarms.

Further sophisticated statistical tests and tuning of the time series QC are needed to refine the existing RTDMS to a more robust system. Once the Phase II codes have detected a possible problem with data quality, the 24x7 staff needs to have adequate graphics and other tools to decide what may be the real problem and what action if any may be needed.

5. Acknowledgements

The authors are thankful to Mr. Venkat Anne Rao for his contributions to the web-based portion of the work. The authors are grateful to the SDM staff for their feedback on earlier versions of Phase I of the RTDMS. We appreciate the efforts of the NCEP implementation staff for their efforts particularly David Ma. Phase I of the RTDMS was made possible by the earlier operational work on data dump counts by Dennis Keyser of the EMC. Maxine Brown of the NCO has been instrumental in initiating the RTDMS project. Comments on the time series QC problems were useful from Colin Parrett of the British Meteorology Office.

Table 1. The number of occurrences of suspect data categories for the March 2005 test.

Test	SC	SB	SR	VSC	VSB	VSR
Test1	135	123	130	120	36	54
Test2	41	24	31	67	8	14
Test3	11	3	7	54	2	5
Test4	11	3	7	5	2	5

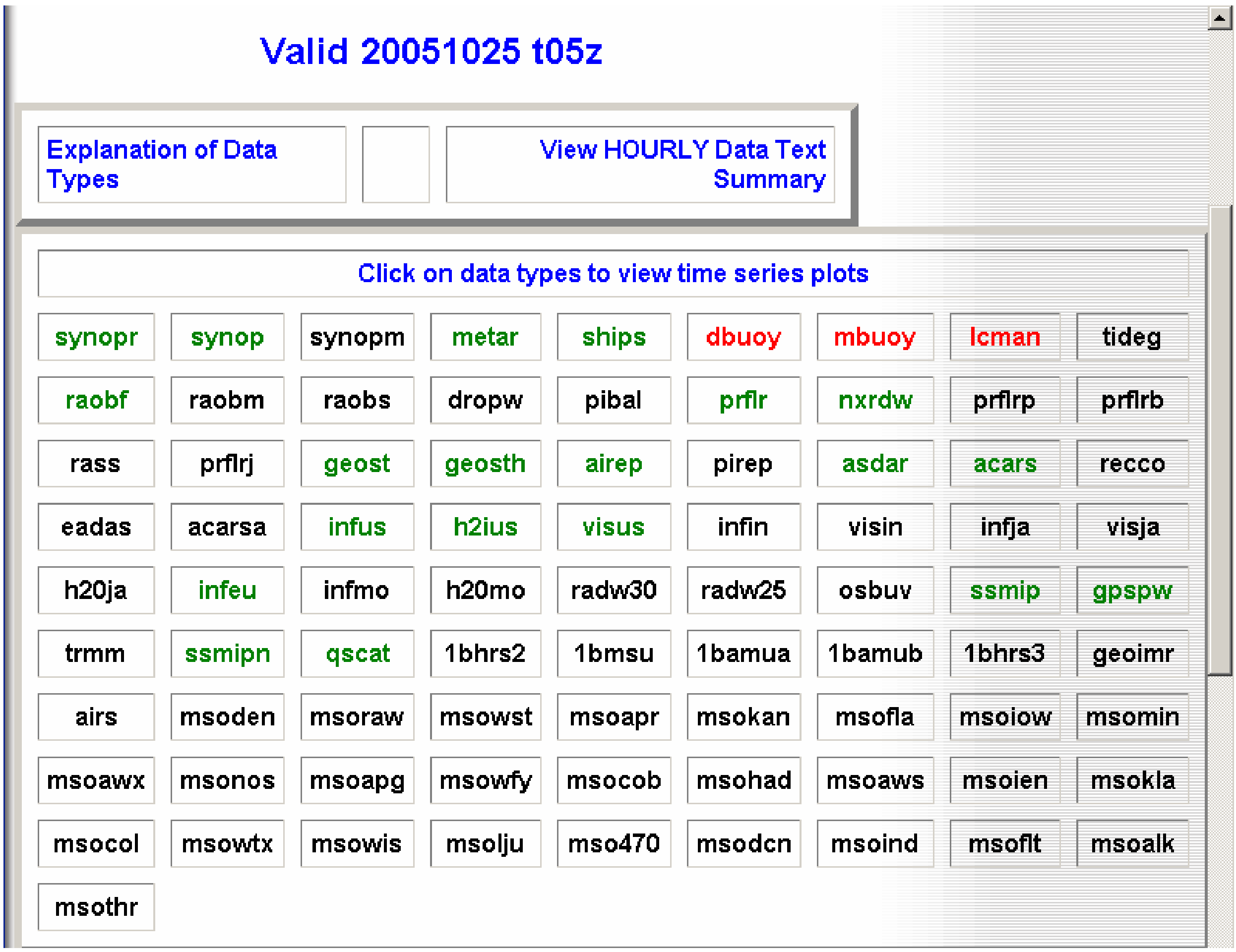

Fig 1. Color Page Summary for Hourly Data 05Z 25 March 2005

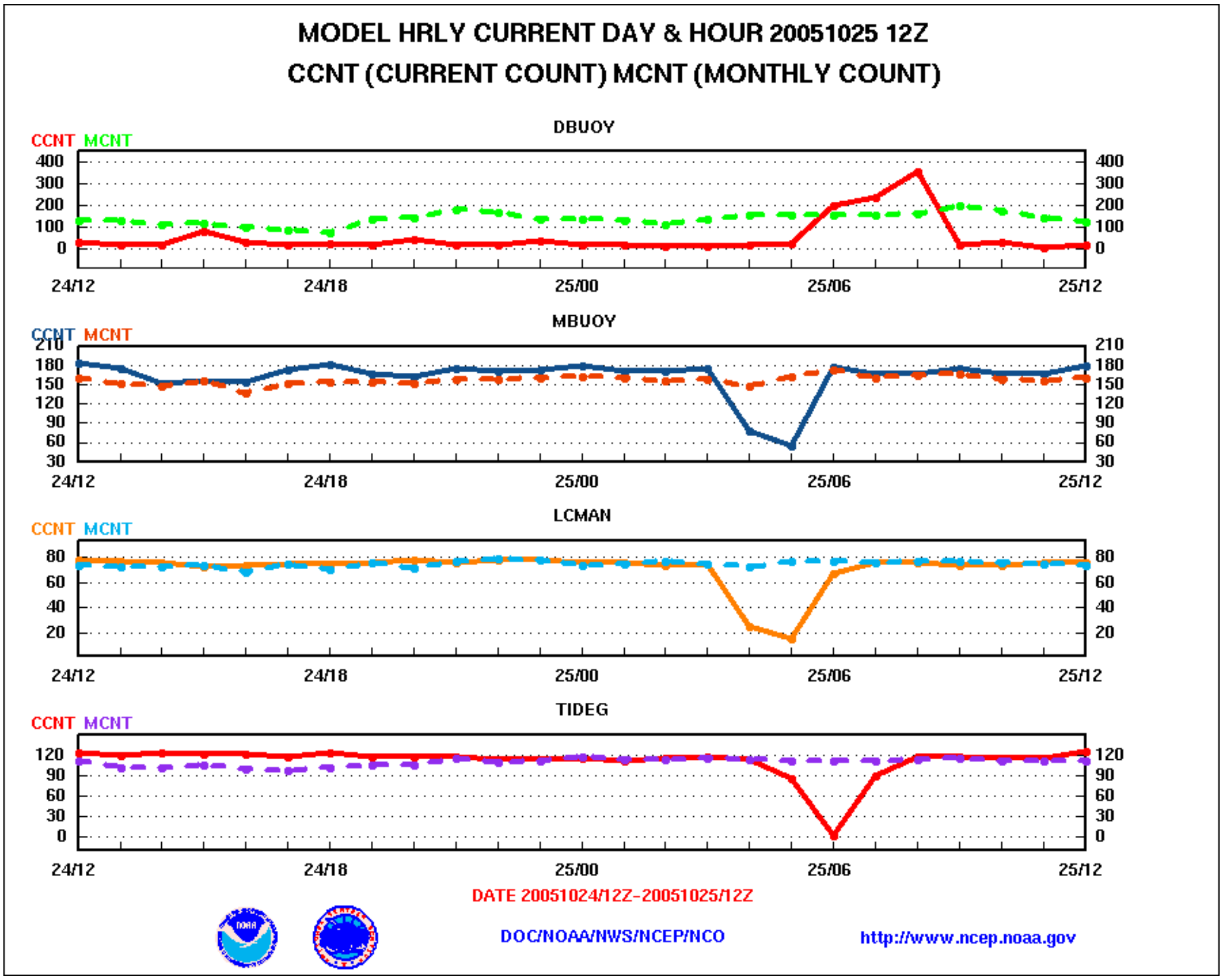

Fig 2. Select Time Series Plots for 12Z 25 October 2005

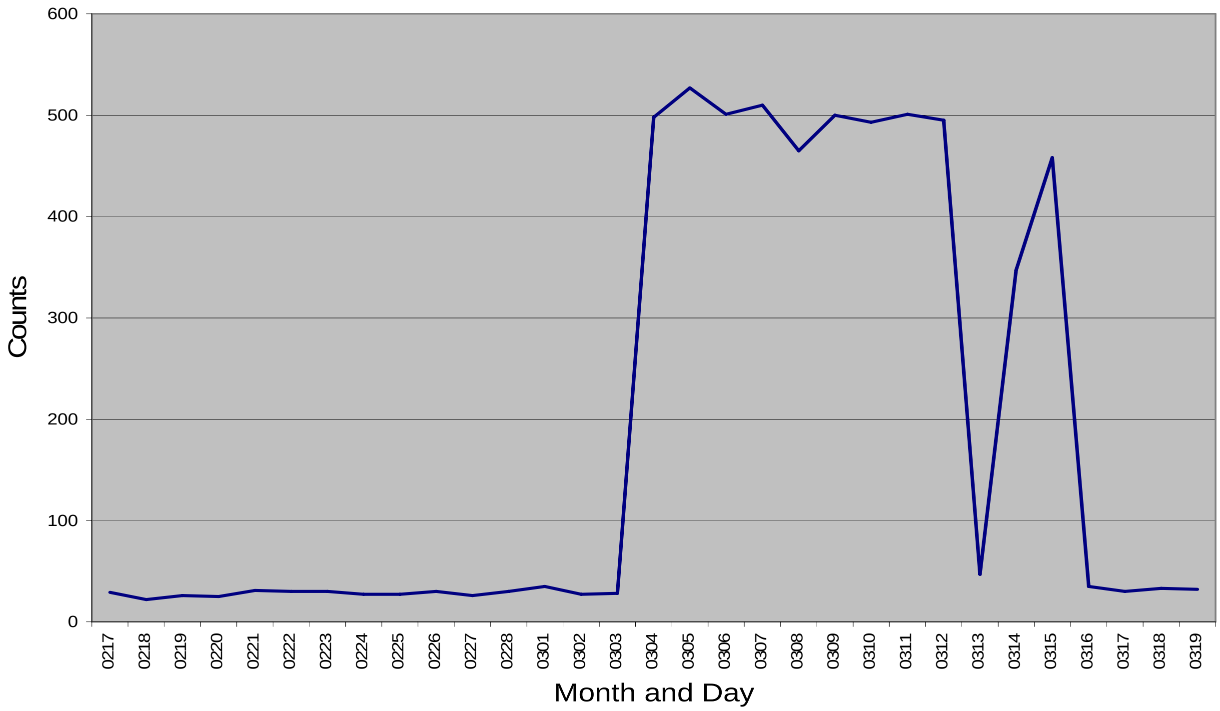

Fig 3. Counts of Marine Virtual Temperature 00Z

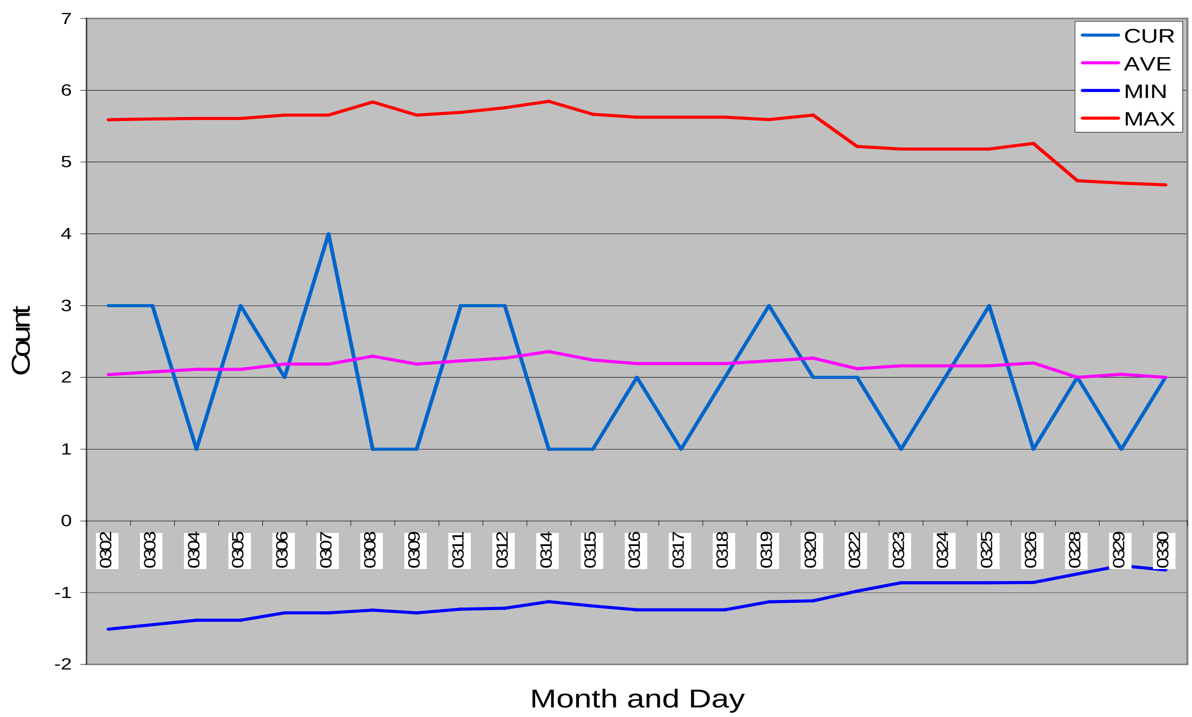

Fig 4. AIREP Temperature Counts Surface to 700 hPa 06Z

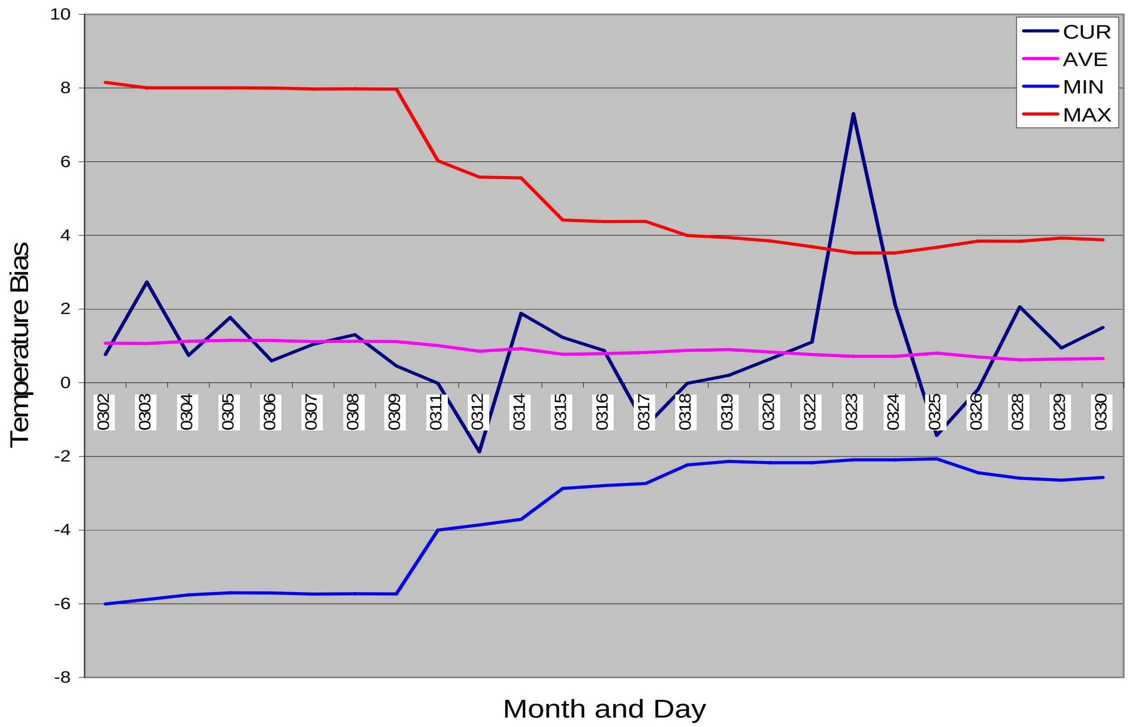

Fig 5. AIREP Temperature Bias Surface to 700 hPa 06Z

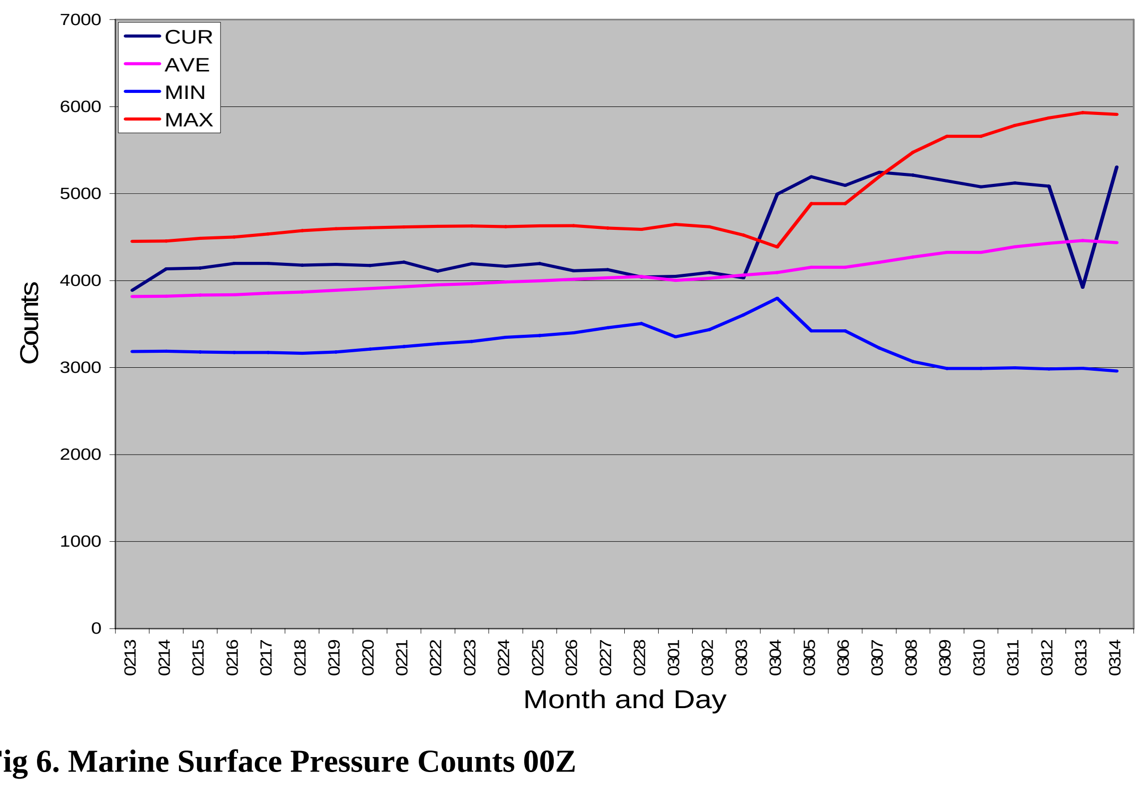

Fig 6. Marine Surface Pressure Counts 00Z

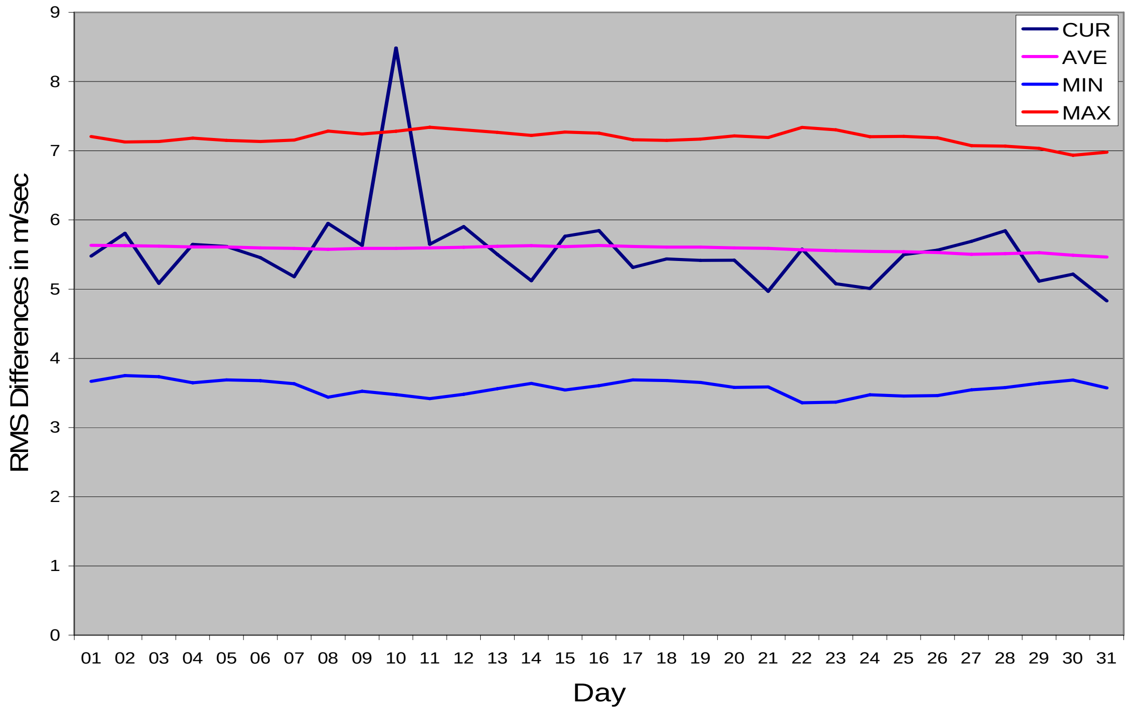

Fig 7. RMS Wind Differences for NESDIS IR Winds 200 hPa and up 00Z March 2005

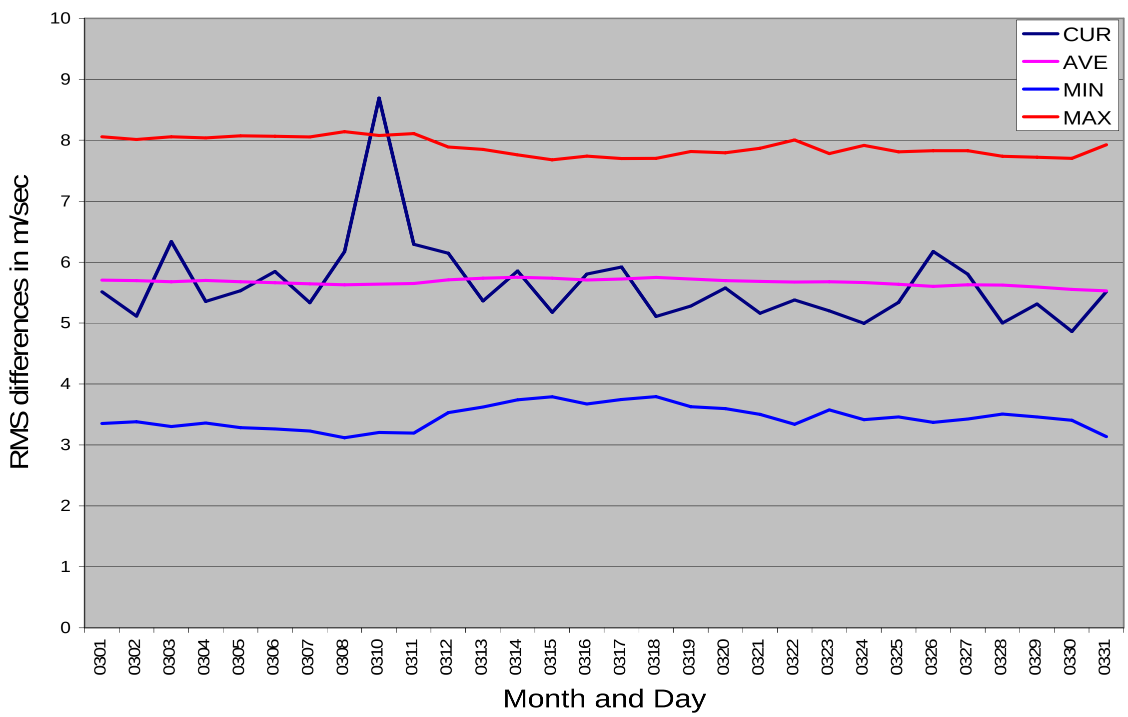

Fig 8. RMS Wind Differences for NESDIS IR Winds 200 hPa and up 18Z March 2005

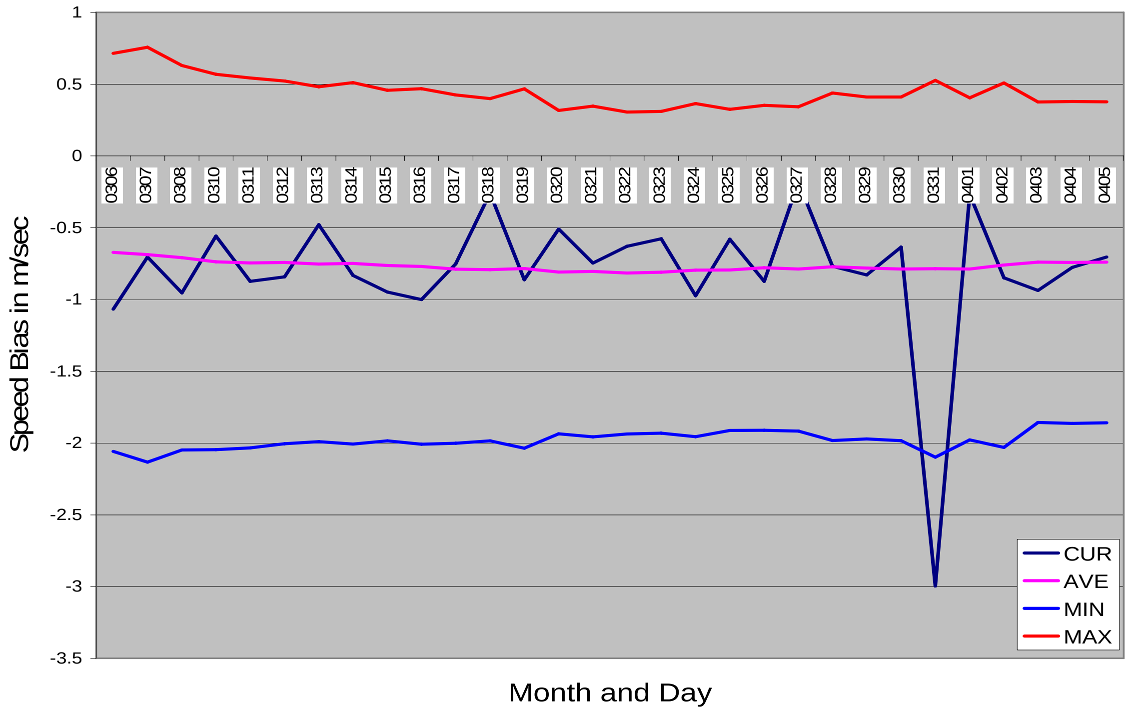

Fig 9. Speed Bias for NESDIS Satellite Winds 700 to 300 hPa 18Z

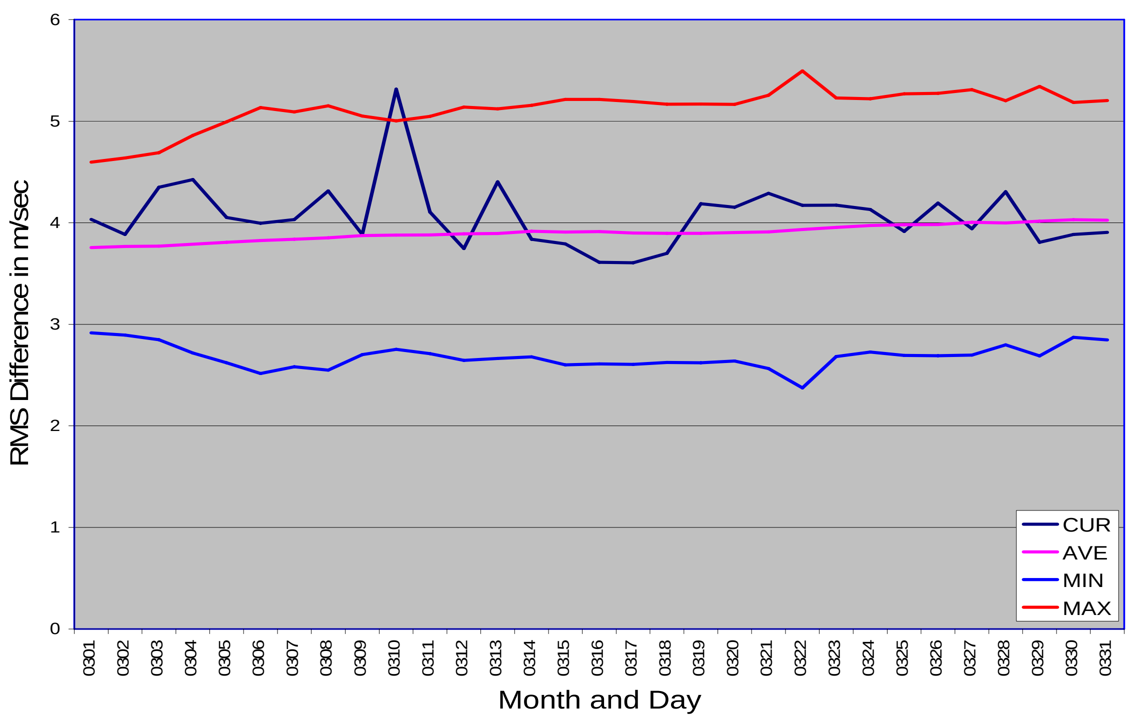

Fig 10. RMS Wind Differences for Satellite Wind Type 257 Surface to 700 hPa 06Z

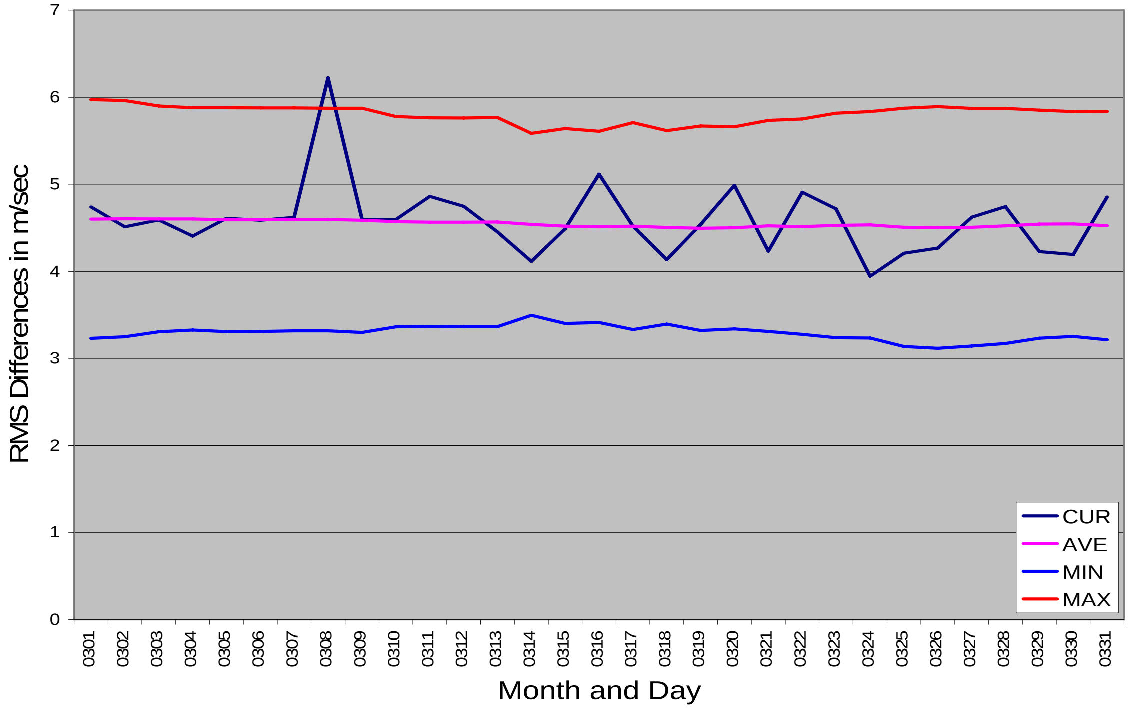

Fig 11. RMS Wind Differences to Guess for Marine Data Type 280 18Z

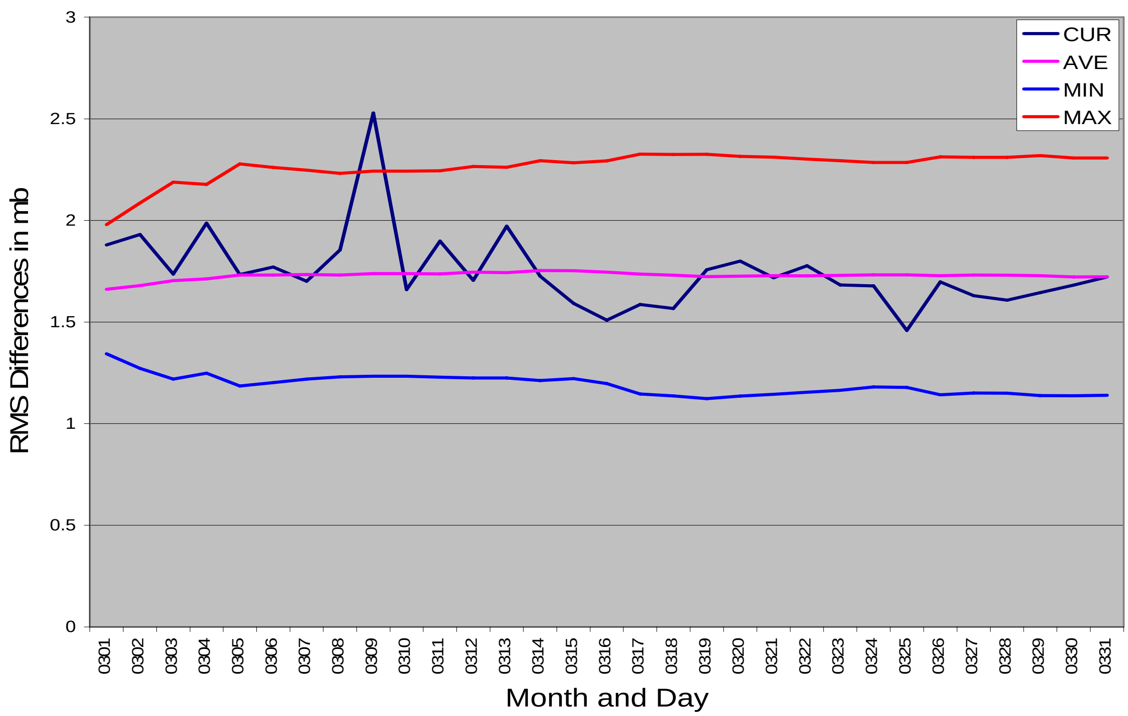

Fig 12. RMS Surface Pressure Differences to Guess for Marine Data Type 180 00Z