JP1.30 QUALITY CONTROL OF PILOT BALLOON NETWORK DATA FOR CLIMATE MONITORING

Javier Murillo[1], John Mejia, Jose Galvez, Raquel Orozco
CIMMS/University of Oklahoma
Norman, Oklahoma

Michael Douglas
National Severe Storms Laboratory
Norman, Oklahoma

1. INTRODUCTION

[1]Special observations have been made since 1997 to monitor the variability of the wind field over the intertropical Americas with a project known as the Pan-American Climate Studies Sounding Network (PACS-SONET), supported by the PACS program (now part of the Climate Prediction Program for the Americas, CPPA) of NOAA's Office of Global Programs. These observations consist of once or twice-daily pilot balloon observations made at up to 20 locations, currently in 8 countries ranging from Mexico to Paraguay. Since the database now consists of more than 40,000 observations and is sufficiently long to be starting to be useful for evaluations of short-term climate variations and verification of reanalysis products, the quality control of the data set is of special concern.

This presentation will describe the procedures we have developed to produce the research quality data set now being prepared for more widespread distribution. These steps include the use of special software at the observation site to reduce the inevitable observer errors, the plotting of the data and calculation of means to detect possible errors, and the comparison of nearby radiosonde observations for detection of possible systematic errors. The frequency of erroneous data and the sources of these errors will be discussed, as well as their impact on the usefulness of the overall data set. Finally, we describe future activities related to maintaining the quality of the growing database. Some of procedures described here can be applied to other pilot balloon data sets that continue to be generated in different parts of the world.

The PACS-SONET database

Since the initiation of the PACS-SONET project in 1997, more than 75 sites have made pilot balloon observations as part of the project, with approximately 20 active sites being the average number at any one time.

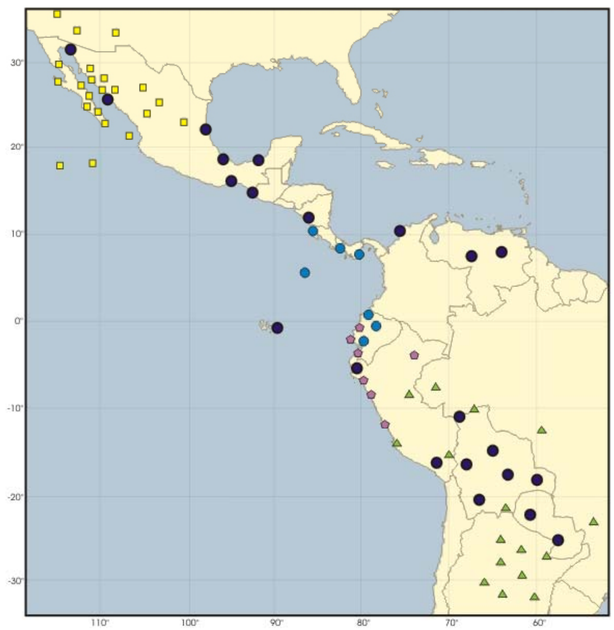

Figure 1. Sites where pilot balloon observations have been made as part of the PACS-SONET project and other allied activities since 1997

To date, the project has generated more than 40,000 soundings. Figure 2 shows the number of observations per month generated by the network from April 1997 through March 2005.

[1] Corresponding author: Javier Murillo, CIMMS, 1313 Halley Circle, Norman, Oklahoma 73069; email: Javier.Murillo@noaa.gov

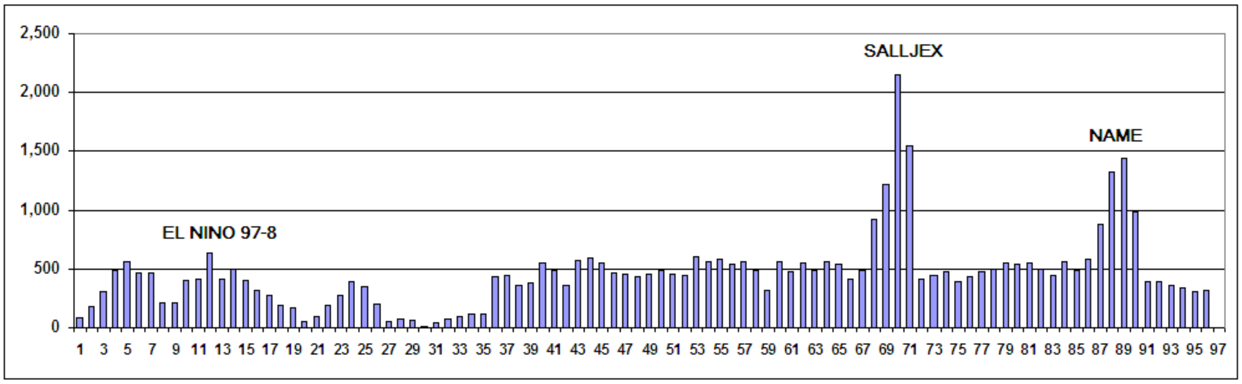

Figure 2. Number of observations per month generated by the PACS-SONET network from its inception in April 1997 through March 2005.

Problems with maintaining data quality fall into many parts. These include:

During the observation stage:

Uncertainties in the balloon ascent rate and deviations from an assumed ascent rate

Errors in reference angle determination and set-up before each observation

Random observer errors (many types)

During the data processing stage:

Failure to use the software to correct the data

Over-smoothing or under-correction of errors

File labeling errors, transmission errors

Data archival stage:

Failure to detect errors in the historical archive

Inadequate metadata to document all changes in procedures

2. ERRORS DURING THE OBSERVATION STAGE

Uncertainties in the balloon ascent rate—deviations from an assumed ascent rate

The calculation of the wind from measurements of the elevation and azimuth angle of a pilot balloon assumes knowledge of the height of the balloon. This ascent rate is assumed to be a constant, independent of height. The basis for this assumption is a series of double theodolites measurements that have been made to determine directly the height of the balloon as a function of time. Such measurements show scatter in the linear ascent rates of about 5% about a mean value. This uncertainty is present in every pilot balloon sounding and cannot be reduced easily. Slight imperfections in each balloon, small inflation errors, and the possibility of small vertical motions in the atmosphere all contribute to this uncertainty. Atmospheric vertical motions are generally small above the boundary layer and away from possible shear instability generated turbulence. However, within a windy boundary layer the turbulent vertical velocities could be a significant fraction of the calculated ascent rates we have found for our pilot balloons (~3.2 to 3.8 m/s, depending on the weight used for the nozzle).

To minimize the ascent rate uncertainties we have several options. To eliminate the ascent rate mean value uncertainty we make enough double theodolites measurements to assure that the means are statistically stable. This has been done to a certain extent, thought there are many soundings required. This is because there are two gas options, helium and hydrogen, which produce slightly different ascent rates for the same balloon lift, and because the weights have been changed more than once during SONET to save gas, in locations where cylinder gas was very expensive and hard to obtain locally.

To reduce the ascent rate uncertainty due to atmospheric turbulence most pilot balloon sites launch in the morning, close to 1200 UTC, which for most sites is between 0500 and 0800 LT. A secondary reason is that low cloudiness is less at this hour at most sites. At some sites 1200 UTC is dark, and in this case the observations are made about 30 minutes after sunrise. The advantage of a launch time that is relative to sunrise is that the boundary layer is stably stratified and much less subject to major ascent rate uncertainties due to dry convection that occurs during the late morning and afternoon over land areas. Afternoon observations, where needed for operational or cloudiness reasons (too much AM cloudiness) are made late in the afternoon (ideally near 0000 UTC), when the boundary layer has begun to stabilize to dry convective motions.

In order to further verify the ascent rate used in pilot balloon soundings, averages can be calculated from comparable data sets and the results examined for consistency. The opportunity arose during the North American Monsoon Experiment (NAME) that took place in the summer of 2004. During NAME, special pilot balloon observations were made twice a day in Empalme (27.95° N), which is also a regular radiosonde site. Separate means were obtained for the month of July using both wind data from pilot balloon soundings and GPS-derived winds from radiosonde observations. The results, shown in Figure 3, suggest that no major shift of the assumed ascent rate exists for pibals, as u and v curves follow closely the radiosonde data and heights at which maxima and minima occur are essentially the same for both data sets.

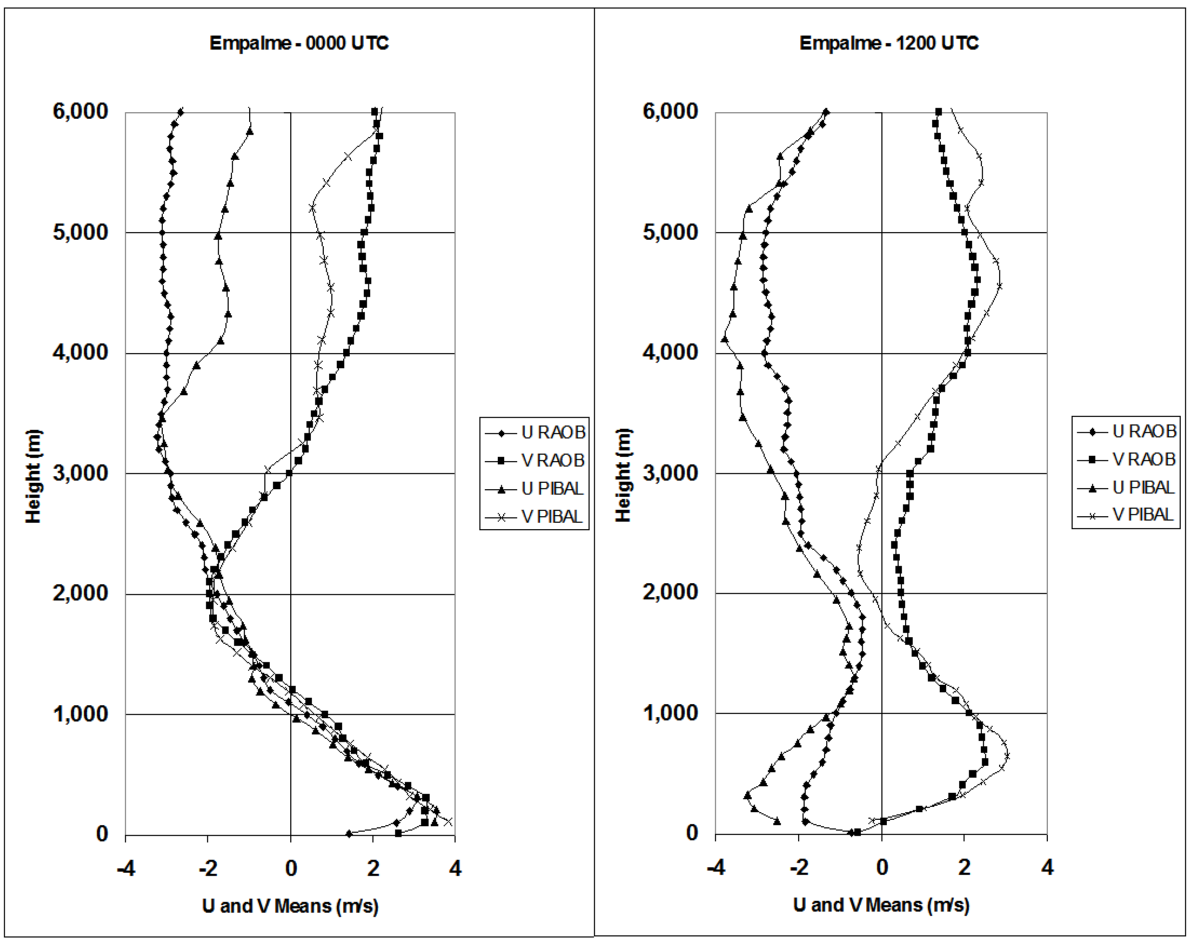

Figure 3. Means of U and V from pilot balloon winds and from GPS-derived radiosonde winds. The left panel shows means for July, 2004, using 0000 UTC data only and the right panel shows the same except for 1200 UTC. Reasonable agreement exists between the pilot balloon and the radiosonde mean profiles, which suggests that, in the mean, the assumed ascent rate for pilot balloons is close to the actual rise rate.

Errors in reference angle determination and set-up before each observation

An error that can have a major impact on the mean winds at an observing site is the incorrect determination of the reference angle. Before each observation the theodolites must be adjusted so that it reads 360 degrees when pointed north (the wind calculation program assumes this). If the true north was not established correctly at the site every observation will have this systematic error. For this reason we prefer to establish the reference angles ourselves during an initial site visit. However, the site can be moved for logistical reasons and we have not been able to visit a few sites.

To reduce north-determination errors at a site, our web page details finding north via sun position, or by using a hand-held GPS. We prefer both methods be used to determine north to better than 1 degree. A systematic error of this magnitude should be acceptable. Unfortunately, some individuals may not have portable GPS's, or may not have access to internet web sites that provide astronomical positions for any location or time.

Another source of reference angle error comes from the design of the theodolites, and the fact that the azimuth angle lock screw can be left un-tightened, and thus capable of rotating during the sounding. This happened at one site during SONET on a daily basis and resulted in 4 months of data being lost (no easy correction of the azimuth angle is possible in this case).

Random observer errors

There are many possible random errors that observers can make during the pilot balloon observation. The most common type is obviously the angle reading error. Elevation and azimuth angles are to be read every minute (every 30 seconds during the first 8 minutes of the observation) and it has to be done quickly enough so the balloon don't get lost from sight. Yet the readings have to be accurate, ideally to 0.05 of a degree or better. An observer with experience will typically make a few errors of this type during a 50-minute sounding. These errors are easily removed from the observation with the use of software especially written to process the angle data. We have found from our experience that most errors are not in the decimal part, but in the integer part of the angle read. Errors of 1, 5, 10 and even 100 degrees are far more common than fractional errors.

A special case of the above type of error is generated when the scale is read in the wrong direction. This occurs often to new observers due to the fact that the theodolite commonly used by our project has an azimuth scale that runs from right to left. This type of error and the one mentioned previously can be easily corrected using available software, as long as there are few bad readings among many good ones.

There are other types of errors which are random in nature and are due mainly to bad observing practices or lack of attention from the observer. These may occur when the angles are not read at the right time (on the top of the 30 or 60-second interval), when the reading is taken while the operator is still moving the theodolite controls or while the theodolite is not pointing at the balloon (balloon not in the center of the cross-hair reticule). These errors are usually small (a few tenths of a degree), but can be significant especially when the wind being measured is strong in magnitude.

Missing readings

One or more readings can be missed during a pilot balloon sounding due to an obscuring object such as cloud (by far the most common), fog, trees, etc. Another frequent reason for missing readings is the observer's lack of ability to track the balloon due to its relatively fast movement or an overhead position. Often times the observers are able to resume the observation after the balloon has moved away from the obscuring object. When one or more readings are missed, current software fails to adjust for a longer interval between points and generates wrong wind values at the beginning and at the end of the gap. As a first approach to solving this flaw, the wind values at the affected levels would be removed from the observation, but a better solution has been developed that consists of a program that interpolates the balloon's position and generates wind values for short gaps of 1 or 2 minutes. This is illustrated in Fig. 3.

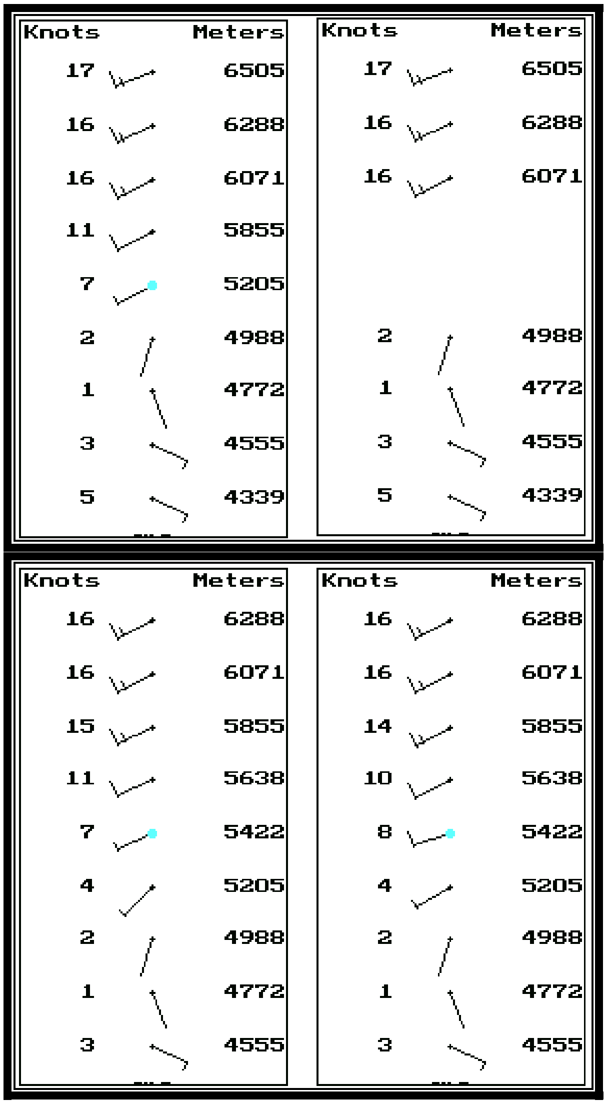

Figure 3. The upper panel shows vertical wind profiles in which data points have been artificially removed to simulate missing data (left). The original profile along with other obtained by linear interpolation of the balloon angles is shown on the lower panel.

3. ERRORS DURING THE DATA PROCESSING STAGE

Throughout the SONET project, software to process the angle data have been made available to observers at the time a new station is established. The software, known as "Corrige", is used to calculate winds and to clean up the data. It consists of a single executable program that reads pairs of theodolite angles, elevation and azimuth, along with the time in minutes these were taken. The program then calculates the wind speed and direction as a function of time. The height of the balloon at any given minute is calculated from the assumed-constant ascent rate. The program also displays graphics of azimuth and elevation vs. time, and a wind-barb profile. A mistaken angle usually stands out in one of the angle plots and in the wind profile. The program allows the operator to use the arrow keys to correct a suspicious angle by adding or subtracting one degree at a time and then to look at the wind profile to see the effect of the correction.

A mistaken angle will generate wrong wind values at two levels of the sounding. The reason for this behavior is that wind at any given time (or height) is calculated using the balloon's positions observed on the minute previous to and on the minute following the time being considered. This way, the wind so calculated will be the average of a layer centered at that height. Every angle measured (except for the one at the end of the sounding) is used twice in the wind calculations and this is why a wrong angle generates wrong winds at two levels (Refer to Fig. 3).

"Corrige" is good in many regards, but it lacks a module for entering data. The theodolite angle data have to be typed in using a separate application (any program with the capability of storing plain text files can be used). This fact is the source of a number of errors, some of which, unlike reading errors described in Part 1, are not easily detected by the Corrige program. Some of them are relatively common and include the following.

Input data file errors

The format of the input file required by "Corrige" is quite simple. Two heading lines are required, so angle data must start on the third line and be typed in three columns separated by one or more spaces or tabs. The first column will be the time in minutes, the second and third columns will be the elevation and azimuth angles. If the two-line heading is not present, the program will still skip two lines so it won't read in two lines of actual data. In a few other cases, the elevation and azimuth angles may have been mistakenly exchanged. Since an exchange of columns will result in outrageously bad data, it can be easily detected and corrected.

Decimal point has to be used in the input file. If comma is used instead of decimal point, any fractions are discarded by the program. If a decimal point is missing or misplaced, it will usually result in time or angle values which are too large. These errors are hard to detect and correct with just the use of the Corrige software, so a simple program was developed that reads in the raw data files and checks for and replaces commas or other characters that may have been mistakenly entered.

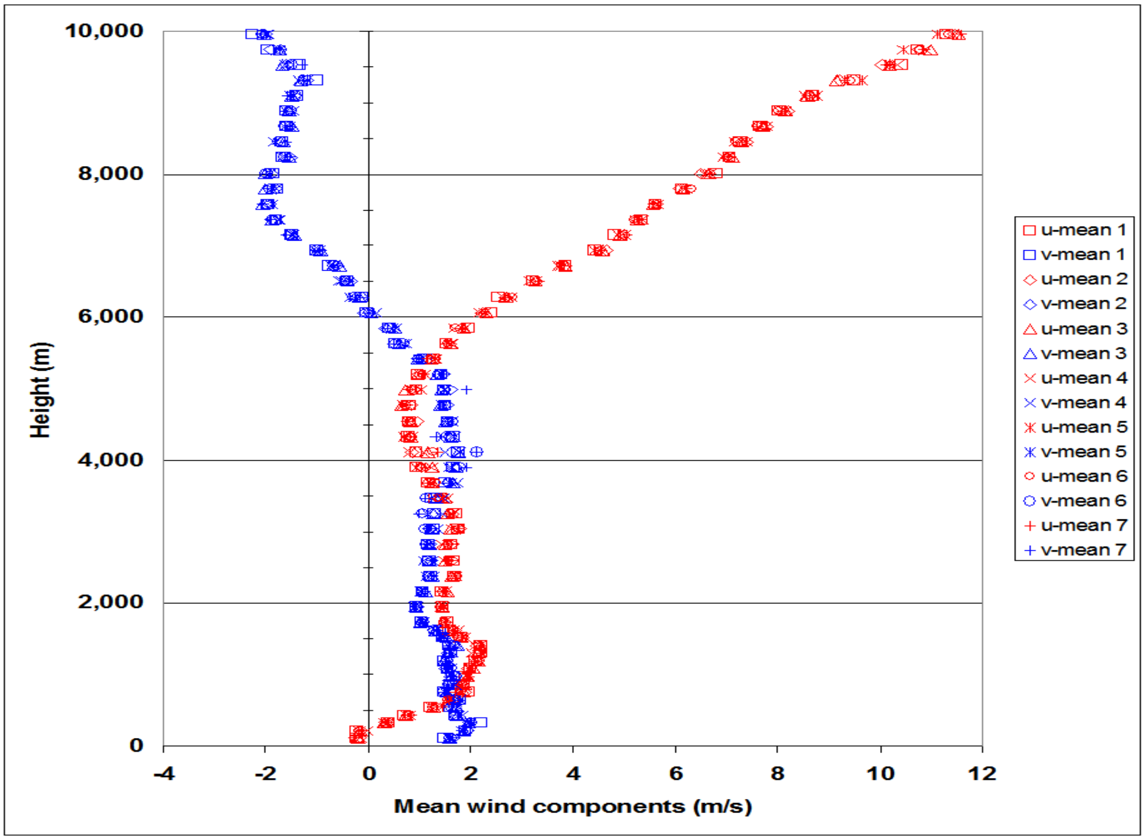

Figure 4. Scatterplot of mean u and v wind components obtained from the same set of 50 observations processed by 6 different individuals. Figure shows that no significant differences can be found in the means. Differences are less than 1 m/s at any given height, and in the lowest 6 km of the mean sounding the differences from the mean rarely exceeded 0.3 m/s (not visible on this plot).

The process of assigning a name to the input data file is related to the data-entry stage and another possible source of error. The filename is limited by the program to 8 characters plus a 3-character extension. The filename is chosen so it can be used later to identify the location and the time the observation was taken; it normally includes the station's identifier, the month, the day and the hour; the latter is codified with a single-digit indicator for the time of the observation: figures 1-8 are used, one for each of the main synoptic hours. Since the file name is constructed manually, this is another potential source of error. This does not affect the data quality directly, but does impact their availability in different ways, especially if they need to be disseminated in real time.

The weather service of Paraguay has contributed to the project with developing a data entry program that prevents many of the above errors from being committed. However, the software cannot be used in every situation since features like the list of station identifiers, the file naming scheme and the maximum length of the sounding are all fixed. The program, designed as a Corrige companion, is being utilized in most SONET stations and has been very helpful in avoiding many typographical and formatting errors.

Errors processing the data

“Corrige” software is an essential tool for removing the most common observational errors in a pilot balloon sounding. In all but one or two stations, it has been made available to the observers so they are actually able to see the wind profiles and not only the raw data. This way, it is hoped, they will learn from their errors and will remain motivated to make good observations. The type of errors that are most common during this stage derive actually from the fact that often times the observers don’t spend enough time correcting the observation in their rush to process and transmit the data. Errors may arise as well at the time the observer enters the station’s altitude and the ascent rate into the program. Additionally, errors not always stand out in the plots displayed by Corrige. If they occur when the angles are naturally changing, then it will be hard for an inexperienced observer to catch them. Skilled observers, however, should be able to discover any errors by looking alternatively at the wind barb profile and at the elevation and azimuth angle plots.

Over-smoothing or under-correction of errors

Due to the subjective part of processing data with Corrige, there might be concerns that large differences may exist when results obtained by different observers, working on the same data set, are compared. It is reasonable to expect some observers to over-smooth legitimate variations, while others will fail to correct existing errors. On an attempt to address these concerns, an exercise was devised in which 7 persons processed the same set of 50 observations using Corrige. Means of the wind components, u and v, were obtained for each individual’s set of corrections and the results compared (Fig. 4). In general, a carefully-made pilot balloon sounding will have little observational errors and these will be easily corrected with no uncertainty in the validity of the corrections. Subjectivity arises only when the observations are made by less skilled observers.

3. QUALITY CONTROL OF THE HISTORICAL ARCHIVE

We are aware of the many observational and other types of errors that can occur in pilot balloon observations mainly due to the fact that these observations are monitored and disseminated in real time. Synoptic maps are generated in our web site, presenting wind data from both the PACS-SONET pilot balloon network and from radiosonde observations. A quick assessment is made of the quality of the pilot balloon data by examining these maps as well as the individual observations. In case we suspect of problems with the pilot balloon data, the observers are contacted and feedback is obtained until the source of the problem is identified.

Our project does not have enough resources to maintain a permanent monitoring of the data being transmitted and we recognize that errors exist in the historical database. On an effort to improve the quality of the data archive, every sounding is being passed through “Corrige”. As it was explained in the preceding paragraphs, this software is interactive and requires an operator to check and correct obvious errors. When we started this effort it was expected that within a timeframe of a few months the process would be completed. However, the database has been growing at a rate faster than expected and the first clean version of the database might not be very close to be released if we continue to use this approach. Instead of applying “Corrige” to every sounding, we have developed a procedure to visually check the data and create a list of files that need to be “corriged”. This procedure, unlike “Corrige” itself, does not require going through many steps to revise an observation. With this procedure we expect to significantly speed up the clean up of the data since it involves the reprocessing of only selected data only and not of the complete archive.

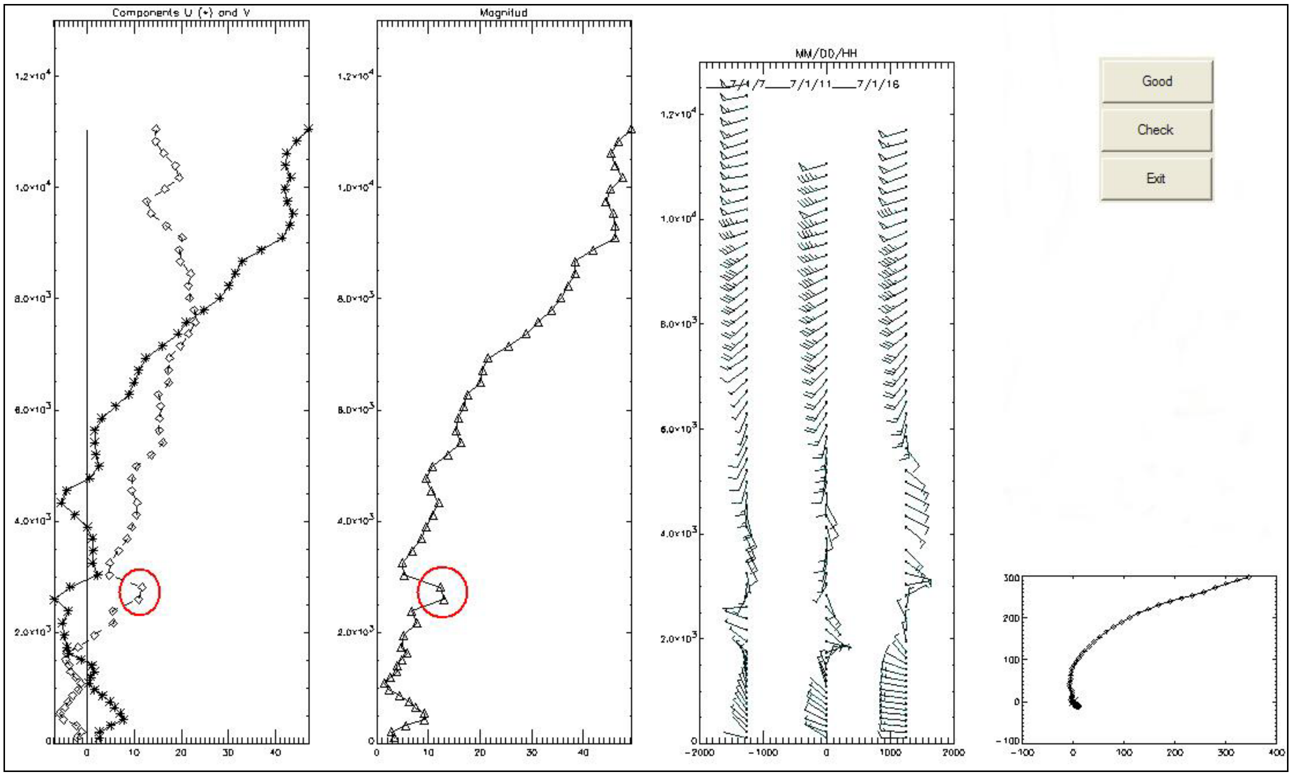

Figure 5. Screen-shot of the program being used to quickly identify soundings that may need further corrections.

Conclusions

Software with the capability of automatically (as opposed to interactively) correcting errors in the observations is being developed in order to speed up the quality control of the PACS-SONET database. The software currently available has little or no built-in error checking capabilities. It was designed to visually inspect the input angles, identify obvious errors and interactively make corrections. The type of errors that the software was designed to detect would be as well detected and corrected using appropriate automated methods. Running means, median filtering, vertical shear calculations or some combination of these, can be implemented in software to be developed. Although these methods are approximate solutions compared to the manual method (supposed to aim for an “exact” solution), automation of this part of the quality control process will result in enormous time savings and will significantly increase the confidence in the data as the methods used and even individual corrections made to the observations could be documented as metadata. Additionally, the wind profiles obtained after the data has been cleaned up through the “approximate” methods would not differ significantly from those obtained using the current method. Errors that the current software was not designed to neither detect nor correct will be easily detected and removed with new software (mainly typing and formatting errors in the input data files). Some results of this effort can be seen in the PACS/SONET web page under a new section named Quality Control.

Current QC methods have been implemented to check for errors in the raw data, i.e. the angle readings via the “Corrige” software. Distance of the balloon from the launching point, wind components u and v, or changes in wind velocity with height can be alternatively checked for errors. When an error has been detected, the correction can be applied easily to the variable from which the error has been discovered or (in theory) to the original angle data.