Strategy to Improve the Homogeneity of Meteorological Data in *Mark3* Databases

K. Le Bail, J. M. Gipson

Abstract Errors in modeling the troposphere is a major part of the error budget in VLBI processing, making the meteorological data very important. Summary tables of missing meteorological data point to the lack of data for important stations in the network (Zelenchukskaya during the CONT08 campaign for example). As a first step in this paper, we study the impact of erroneous meteorological data on the VLBI processing and show that they directly impact the quality of the VLBI solution: the scatter of baseline length WRMS is affected at a significant level (up to 1 mm for 10 years of data), as well as the determination of the Up component: the case of Svetloe during CONT08 shows a linear correlation of 8.9mbar/mm in the case of an offset in pressure. We analyze Westford over the period January 2002 to April 2010, using different sources of meteorological data in the processing: ECMWF, other onsite recorded meteorological data, such as a sensor associated with GPS receivers, and the default value used by *Calc/Solve*. We conclude that using a constant default value in *Calc/Solve* to replace missing meteorological data is not satisfying.

Keywords Meteorological Data, *Calc/Solve*, *Mark3* databases, VLBI processing

1 Introduction

Pressure and temperature data contribute in the VLBI processing in a very significant way. Many studies show the impact of this data via atmospheric modeling, e.g. Davis et al. (1985), Niell (1996), and more recently, Heinkelmann et al. (2009) and Nilsson et al. (2009), to cite only a few. Other studies show the discrepancy found between sensors on the same site and the importance of precise meteorological sensors and homogeneous recording of pressure and temperature, see Niell (2005).

Considering the above studies with great interest, we decided to determine what is in the *Mark3* databases. We constructed yearly tables gathering all the information on pressure and temperature. An example is given in Figure 1 for the year 2008. The crosses indicate missing meteorological data for the stations observing during the session, while the large dots indicate the presence of meteorological data in the database. The small dots indicate the station does not participate in the session. For 2008, the CONT08 campaign stands out: only eleven stations are participating continuously. In this figure, we also see that for some stations, like Fortaleza and Zelenchukskaya, two major stations in the network, more than 90% of the meteorological data is missing in 2008. In the case of Westford, the meteorological data are provided in the past years, by a sensor from the SuoMinet network (an international network of GPS receivers, *http://www.unidata.ucar.edu/data/suominet/*). The absence of data during the last trimester of 2008 show that those data have not been uploaded. When looking at the other years, the same remarks are still valid: some of the major stations are missing a large amount of meteorologic data, other stations change the source of meteorologic data. To emphasize the impact of such discrepancies, we focused on different periods. First we analyzed the eleven stations of the CONT08 campaign, and we show results for Svetloe

Karine Le Bail and John M. Gipson
NVI Inc / NASA Goddard Space Flight Center, Greenbelt, MD 20770, U.-S.

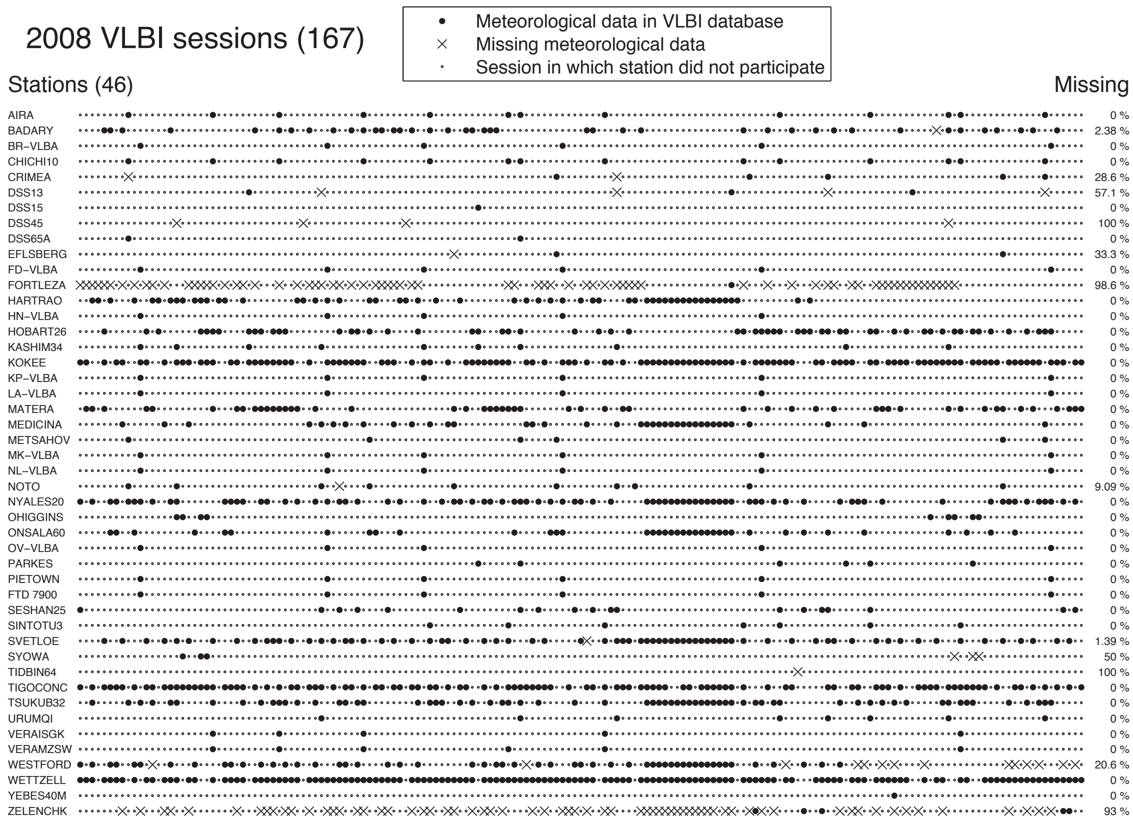

Fig. 1 Summary table of the presence of meteorological data in the *Mark3* databases for the year 2008

and Zelenchukskaya. We also studied Westford from January 2002 to April 2010, considering only the R1 and R4 sessions. The results obtained are presented in section 2 of this paper. In section 3 we discuss a strategy to obtain a homogeneous data set and give conclusions of this study.

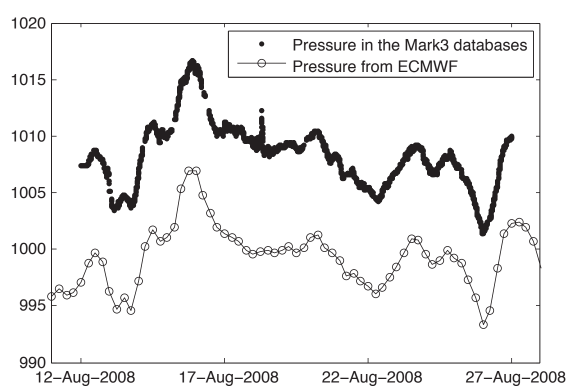

Fig. 2 Pressure in mbar for the station Svetloe during the campaign CONT08, as in *Mark3* databases and as obtained using ECMWF model

2 Impact of erroneous meteorological data in the VLBI processing

2.1 *Comparison with ECMWF data and pressure offset: the case of Svetloe*

Figure 2 compares the pressure found in the database with pressure from the ECMWF model (derived at the Vienna Institute for use with the VMF, Boehm and Schuh (2004)) for Svetloe. There is an offset of about 10mbar over the CONT08 period. To study the effect of this offset, we performed different simulations, subtracting a constant value (2mbar, 5mbar and 10mbar) from the pressure in the database for only Svetloe, and then we solved for station positions over the CONT08 period for the eleven stations. In Figure 3, we plot the differences in U, E, N determination using the pres-

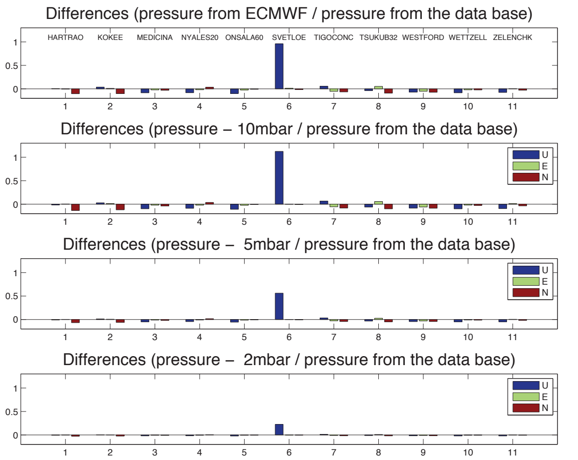

Fig. 3 Changes in position of the eleven CONT08 stations between using for Svetloe either the *Mark3* databases pressure or a modified pressure (ECMWF or biased value)

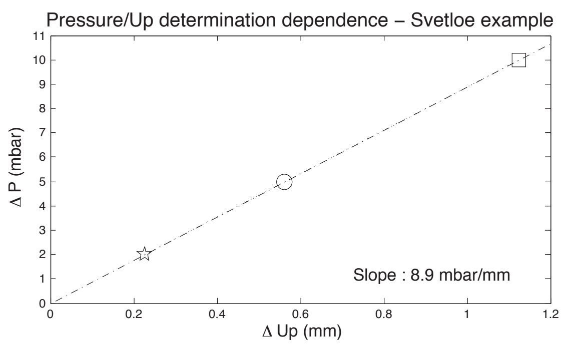

Fig. 4 Dependence in pressure and Up determination for the station Svetloe

sure from the database wrt a modified pressure: 1). the ECMWF pressure, 2). the database pressure minus 10mbar, 3). minus 5mbar, 4). minus 2mbar. The major effect is on the Up component determination of Svetloe. The Figure 4 plots the differences of the Up component of Svetloe versus the difference in pressure (10mb, 5mbar and 2mbar). A straight line is estimated and perfectly fits the three points: the slope has a value of 8.9mbar/mm.

2.2 Default value in *Calc/Solve*

In the case of missing data, the strategy adopted in *Calc/Solve* is to use a fixed and constant value. To quantify the effect, we studied two examples: Zelenchukskaya over a period of two weeks (CONT08) and Westford over a period of nine and a half years.

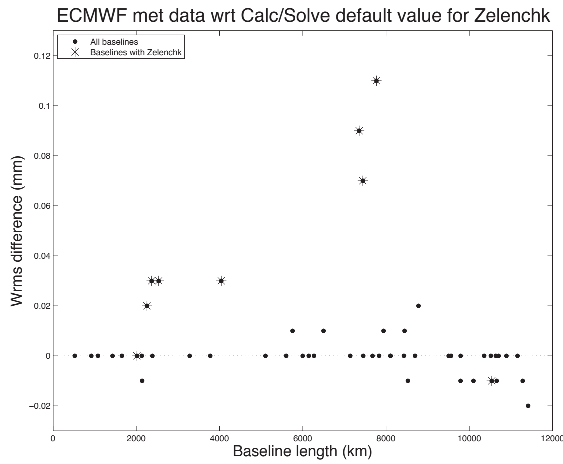

Fig. 5 Differences in WRMS between using for Zelenchukskaya the *Calc/Solve* default value pressure and using ECMWF pressure, in function of the baseline lengths during CONT08. The baselines with Zelenchukskaya are indicated with a star

2.2.1 The case of Zelenchukskaya during the CONT08 campaign

Zelenchukskaya is studied over the CONT08 period. We compute solutions using in one case the *Calc/Solve* value, and, in the other, the ECMWF value. Figures 5 and 6 show the impact of this strategy. In Figure 5 we plot the difference in WRMS of the scatter in baseline lengths as a function of baseline length. The points in the upper half of the plot show a reduction in WRMS in using the ECMWF meteorological data rather than the default value. This improvement in WRMS is up to 0.12mm for some of the baselines. Figure 6 is the graph of differences in the Up component determination affecting the eleven stations of CONT08. The impact is mostly on the Zelenchukskaya Up component with a displacement of 0.31mm.

2.2.2 The case of Westford over the period January 2002 to April 2010

To study Westford, we use data from 351 R1 and R4 sessions from January 2002 to April 2010. The pressure and temperature are collected from 1). The *Mark3*

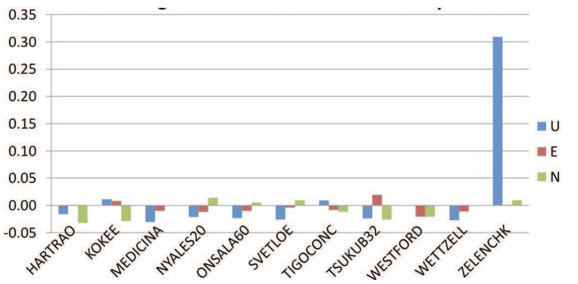

Fig. 6 Effect on using ECMWF meteorological data at Zelenchukskaya instead of the *Calc/Solve* default value. Changes in position of the eleven stations of CONT08

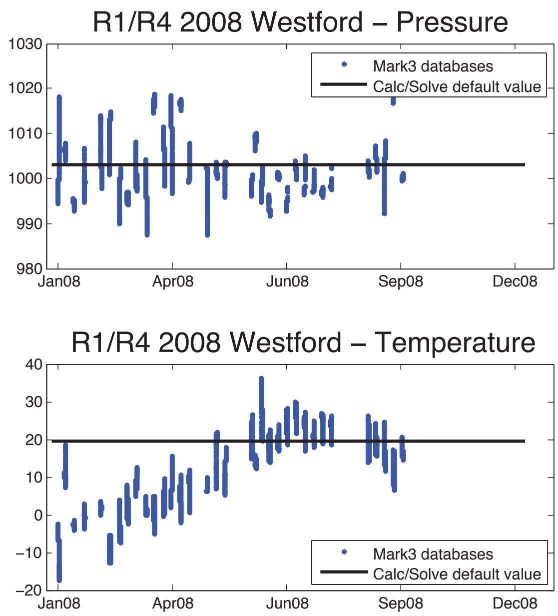

Fig. 7 Pressure and temperature values for Westford in the *Mark3* databases in 2008 and the *Calc/Solve* default value associated

databases, 2). The ECMWF data, and 3). The sensor of the SuoMinet network in Westford (SA01). We compare the results obtained using meteorological data from these three sets and using the default value of *Calc/Solve* when there is no meteorological data in the database. To give an idea of such differences, Figure 7 shows the data from the database and the default value (straight line) in the same plot. It is obvious that using a default value is not realistic. Once again, using meteorological data from the ECMWF or the SuoMinet sensor reduces the WRMS and changes the estimate of Up. In Figure 8, the crosses represent the baselines including Westford and in Figure 9, we plot the differences of

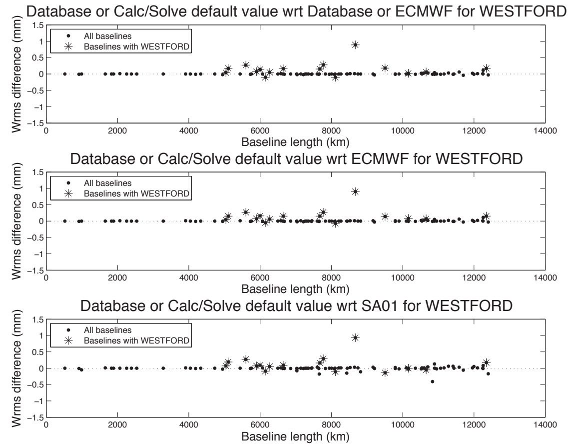

Fig. 8 Differences in WRMS between using different sources of meteorological data for Westford and using the *Mark3* values or the default value. R1 and R4 sessions from January 2002 to April 2010. The baselines with Westford are indicated with a star

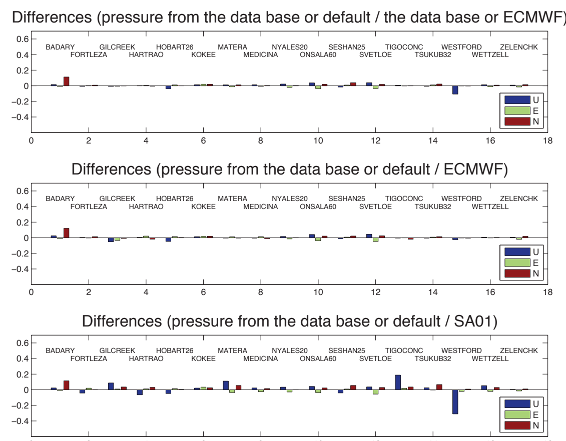

Fig. 9 Changes in VLBI stations position between, for Westford, using either the *Mark3* databases pressure or the *Calc/Solve* default value if no data in the databases, and using another set of pressure (ECMWF or meteorological sensor from the SuoMinet network at Westford)

the Up component. The differences in WRMS reaches up to 0.93mm (Badary - Westford baseline) and affects the Up component of both Badary and Westford by more than 0.1mm.

3 Discussion on a strategy proposed to obtain a homogeneous database and conclusions

In summary, meteorological data in the *Mark3* databases is not homogeneous as it contains missing, biased and inaccurate data. In some cases, the meteorological data in the database comes from another source that has been used to manually fill gaps in the time series (Westford case). This data is not necessarily consistent with existing data. Using the default value in *Calc/Solve* is not a satisfying solution either, as shown in the subsection 2.1.

For those reasons and to achieve continuity in the meteorological data, the data has to be cleaned. Two solutions are then possible.

The first one is to correct the *Mark3* databases to obtain a homogeneous set of meteorological data. This is a long and meticulous work that is currently underway at GSFC/NASA on the existing data. First, this consists of detecting all the bad data (by doing statistics on what is in the database, as well as comparison with other meteorological data) and to correct them with appropriate data, but also to search for accurate meteorological data in the case of missing ones.

The second solution is to have a homogeneous network of meteorological sensors associated to the point of measurement in the global network, observing and recording pressure and temperature continuously. Of course, this solution is not applicable to the existing data and database. But this supports the specification for the VLBI2010 stations, showing the importance of the two parameters pressure and temperature.

References

J. Boehm, and H. Schuh. Vienna Mapping Functions in VLBI analyses. In *Geophys. Res. Lett.*, volume 31, L01603, 2004.

J. L. Davis, T. A. Herring, I. I. Shapiro, A. E. E. Rogers, and G. Elgered. Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. In *Radio Science*, volume 20, pages 1593-1607, November-December 1985.

R. Heinkelmann, J. Boehm, H. Schuh, and V. Tesmer. The Effect of Meteorological Input Data on the VLBI Reference Frames. In H. Drewes, editor, *Geodetic Reference Frames*, International Association of Geodesy Symposia 134, Springer-Verlag Berlin Heidelberg, 2009.

A. Niell. Global mapping functions for the atmosphere delay at radio wavelengths. In *Journal of Geophysical Research*, volume 101(B2), pages 3227-3246, 1996.

A. Niell. SA01 (suominet) - WES2/Westford (GSOS) pressure comparison for 2005 doy 208 (2005/07/27). In *ECGVM MEMO, Massachusetts Institute of Technology, Haystack Observatory*, September 2, 2005.

T. Nilsson, and R. Haas. Impact of atmospheric turbulence on geodetic very long baseline interferometry. In *Journal of Geophysical Research*, volume 115, B03407, 2010.