The VLBI Data Analysis Software νSolve: Development Progress and Plans for the Future

S. Bolotin, K. Baver, J. Gipson, D. Gordon, D. MacMillan

Abstract The program νSolve is a part of the CALC/SOLVE VLBI data analysis system. It is a replacement for interactive SOLVE, the part of CALC/SOLVE that is used for preliminary data analysis of new VLBI sessions. νSolve is completely new software. It is written in C++ and has a modern graphical user interface. In this article we present the capabilities of the software, its current status, and our plans for future development.

Keywords VLBI, data analysis, software

1 Introduction

The currently used software CALC/SOLVE for data analysis of geodetic VLBI observations has been developed since the end the 1970s. It is written mostly in FORTRAN and its design reflects hardware restrictions that existed at the time of its active development phase. The software evolved into a large system that is widely used in many areas of space geodesy. At the same time the system grows stiffer and harder to modify. Continuous commissioning of new VLBI stations, an enormous increase in the observation data rate, and realization of the VLBI Global Observing System (VGOS) technology led to new requirements for data analysis software as well as for the software development process.

Having long experience with the development of the CALC/SOLVE VLBI data analysis software, the VLBI group at the NASA Goddard Space Flight Center initiated creation of the new generation software.

NVI, Inc.

The first step in developing new data analysis software was made in 2007, when the IVS Working Group on VLBI Data Structures (IVS WG4) was created [6]. Efforts undertaken by the group were eventually realized in the creation of the new VLBI data format, vgosDB. In mid-2009 a design of system architecture of νSolve was elaborated and presented at the IVS 2010 General Meeting in Hobart ([2]). A prototype of the software was demonstrated in 2011 at the 20th EVGA Meeting in Bonn [3]. The first version of νSolve was presented at the IVS 2012 General Meeting in Madrid [4]. A public release of the software was made on February 20, 2014.

Currently νSolve is used for routine data analysis of the IVS-R4 and IVS-INT sessions at the NASA GSFC VLBI Analysis Center. It was also used in data analysis of the first VGOS observations.

2 New VLBI Data Analysis Software

νSolve is a replacement for interactive SOLVE. It is designed to analyze a single VLBI session by performing necessary calibrations and data editing and storing results in an appropriate format. Later νSolve will evolve into a powerful session editor that allows us to fix all known anomalies of the VLBI observations, for example subambiguities.

We should note that νSolve does not make global solutions. A separate executable (a driver) will be developed later to perform data analysis of multiple sessions of VLBI observations.

The architecture of νSolve was discussed in detail in [2] and [4]. Here we just outline the main features of the architecture.

The software is written in the *C++* programming language. It is being developed with the Linux/GNU operating system, but its use is not limited only to Linux. We tried to use a minimal set of external libraries for its functionality. In addition to the standard system libraries, *libc* and *libm*, the *Qt* library is used for the graphical user interface, data containers, and auxiliary tools. To realize input and output operations in the new vgosDB format we use the *netCDF* library [8].

The software consists of two parts:

1. The Space Geodesy Library, a library where data structures and algorithms are implemented (about 90% of the total source code);
2. An executable `nuSolve` – a driver that calls the library functions and organizes work with an end-user (about 10% of total source code).

Such organization of the software allows us to reuse the source code in other applications. In the first public releases, while we have only one executable, `nuSolve`, the distribution and the whole software package is called "nuSolve". Later, the library and drivers will be distributed in separate packages.

The modular structure of the software makes it flexible and stable. By a *module* we mean a logical block of the source code that is loosely tied with other parts of the software. A detailed description of modules is given in [2].

3 Features of the Software

The general features of the software are the following. It is able to read and write data in the Mark III DBH format as well as in the new vgosDB format. There are no limitations on the numbers of stations and sources that participate in one session or the number of observations. The software can work either through the CALC/SOLVE catalog subsystem or in a standalone mode. The process of data analysis can be automated to some extent. Special efforts were made to reproduce CALC/SOLVE results. If a user selects the option "Interactive SOLVE compatible mode", νSolve makes a solution that is very close to the solution obtained with interactive SOLVE (the post fit residuals differ by not more than 0.2 ps) using the same set of models, external a priori data, and so on.

3.1 Functionality

The capabilities of νSolve are evolving in accordance with the adopted model of the software production process (see [2]). All necessary key functional points are implemented in the software. νSolve is able to: 1) read/write files in Mark III DBH and vgosDB formats, 2) display on plots various information that were stored in the files, 3) perform least squares parameter estimation, 4) process a single VLBI session and save obtained results, 5) detect and process clock breaks, 6) resolve group delay ambiguities, 7) evaluate ionospheric corrections, 8) calibrate weights of observations to make a normalized χ^2 equal to unity, 9) eliminate outliers, restore previously eliminated observations, 10) use different geophysical models in data analysis, and 11) apply alternative a priori information from external files.

These key functional points have allowed us to use the software in routine data analysis of all IVS-INT and IVS-R4 VLBI sessions since July 2012.

3.2 Plotting Subsystem

A simple plotting subsystem has been developed for νSolve. By design, the subsystem is a separate generic module that is independent of VLBI data structures, estimated parameters, or other parameters. Instead, it provides a mechanism to communicate with other parts of the software. This feature makes it possible to use the plotting subsystem in various aspects of VLBI data processing and other applications.

The use of the plotting subsystem allows a user to display various values that are either stored in files, evaluated "on the fly", or estimated from VLBI data analysis. Also, it permits a user to edit observations by eliminating outliers and manually adjusting the number of group delay ambiguities.

In Figures 1–3, several examples of using the plotting subsystem are shown.

3.3 Estimator

The module Estimator of νSolve performs a least squares estimation of the following parameter types:

- *local parameter*: an unbiased parameter that is determined for a whole session;
- *arc parameter*: an unbiased parameter estimated for an interval specified by the user (e.g., 1 hour);
- *piecewise linear function*: coefficients of a continuous linear function estimated from data, where the interval between nodes is specified by the user;
- *stochastic parameter*: an alternative to a piecewise linear function, a time varying parameter is modeled as a stochasic process.

The realization of the least squares estimation is made with a square-root information filter (SRIF) [1]. Using SRIF and its derivations makes it possible to implement a model where the arc and piecewise linear functions can have different lengths of segment intervals or have overlapping segments.

The software can estimate the following parameters:

- coefficients of the polynomial model for station clocks,
- tropospheric zenith delays and horizontal gradients,
- station positions,
- antenna axis offsets,
- source coordinates,
- polar motion offsets and rates,
- Earth rotation, $d(UT1-UTC)$ and its rate,
- angles of nutation,
- baseline clock offsets, and
- baseline vectors.

The user can assign any of the parameter types to each of these parameters. The user can select a list of stations to estimate their positions or sources to estimate their coordinates. If all available stations or sources are selected, the user can specify what station or source *a priori* coordinates will be used in the equations of No-Net-Rotation and/or No-Net-Translation constraint.

4 Data Processing Operations

The following essential operations that are necessary to perform to make a VLBI session usable in a batch solution are: clock break detection and correction, ambiguity resolution, evaluation of ionospheric correction, adjusting of weights of observations, and outlier processing. We now discuss these operations.

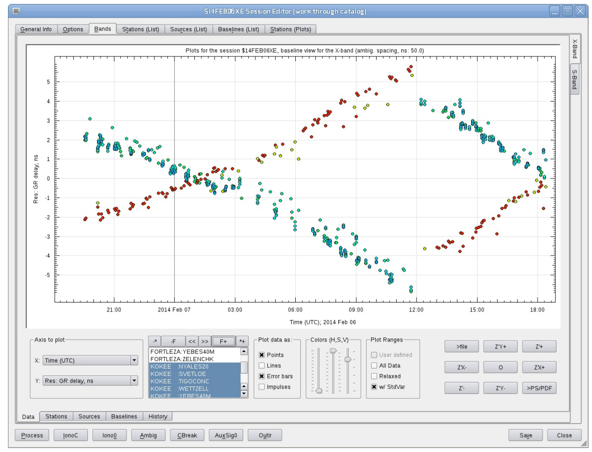

Fig. 1 A clock break at station KOKEE during session \$14FEB06XE. The group delay residuals are shown only for selected baselines.

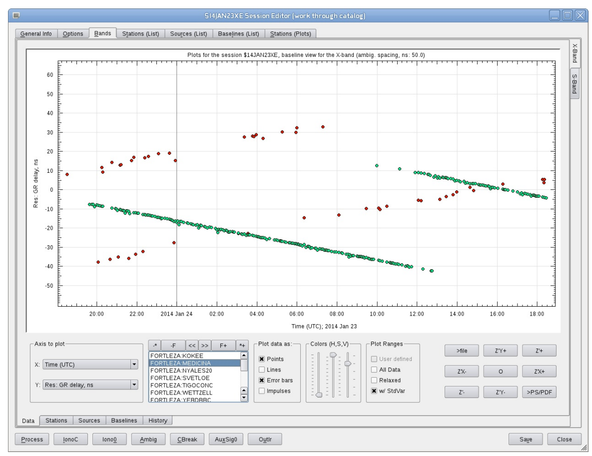

Fig. 2 Group delay residuals with unresolved ambiguities for the FORTLEZA–MEDICINA and MEDICINA–WETTZELL baselines of the \$14JAN23XE session.

A clock break is a discontinuity in the recorded time marks of the observations due to hardware problems at the station. There are also other effects (e.g., manually applied phase calibration which consists of several segments) that manifest themselves as clock breaks as shown in Figure 1. The standard SOLVE software estimates parameters of a clock break as additional parameters of the solution. In contrast, *v*Solve estimates clock break parameters in a separate solution and then applies them in further data analysis. Such an approach allows processing of rare cases of multiple clock breaks during a short interval of time.

Clock breaks can be detected and corrected in automatic, semi-automatic, and manual mode.

Ambiguity resolution of group delays (see Figure 2) is done using the same ideas as implemented in interactive SOLVE. The algorithms implemented in νSolve are less restrictive. The software can process VLBI sessions that have different ambiguity spacing of group delays on different baselines or even on one baseline. In addition, νSolve allows the user to adjust the number of ambiguities manually.

The ionosphere corrections for group delays, phase delay rates, and phase delays are evaluated using dual band VLBI observations. Since the group delays are determined up to an arbitrary number of ambiguity spacings, the evaluated ionospheric correction is not unique. It is a good practice to process clock breaks and resolve group ambiguities before evaluating the ionospheric corrections.

We perform corrections of observation weights to make the normalized χ^2 equal to unity. Complementary standard deviations can be computed in two modes: in a session-wide mode (one weight correction is evaluated for the whole set of observations) and a baseline-dependent mode. Weight corrections change the solution and the distribution of post fit residuals, making the process of weight correction an iterative process. The complementary standard deviations can be imported from an external file. Reweighting is performed in conjunction with the next operation, outlier elimination.

An outlier is an observation with an absolute value of a normalized residual greater than a user-specified threshold. There can be various scales used for normalization of residuals. In the interactive SOLVE compatibility mode νSolve uses a dispersion to normalize the residuals [7]. The dispersion, d is expressed as

$$d = \sqrt{\frac{\sum_{i=1}^{N}(p_i \varepsilon_i)^2}{N-1}},$$

where p_i is a weight of ith observation and ε_i is its residual. The sum is done, depending on options, either over all processed observations or only for a particular baseline. In this case, the normalized residual will be

$$\rho_i = \frac{p_i \varepsilon_i}{d}.$$

As one can see, the normalized residuals are unitless. Distributions of the residuals and the normalized residuals are different. In Figure 3 both types of values are shown for one baseline.

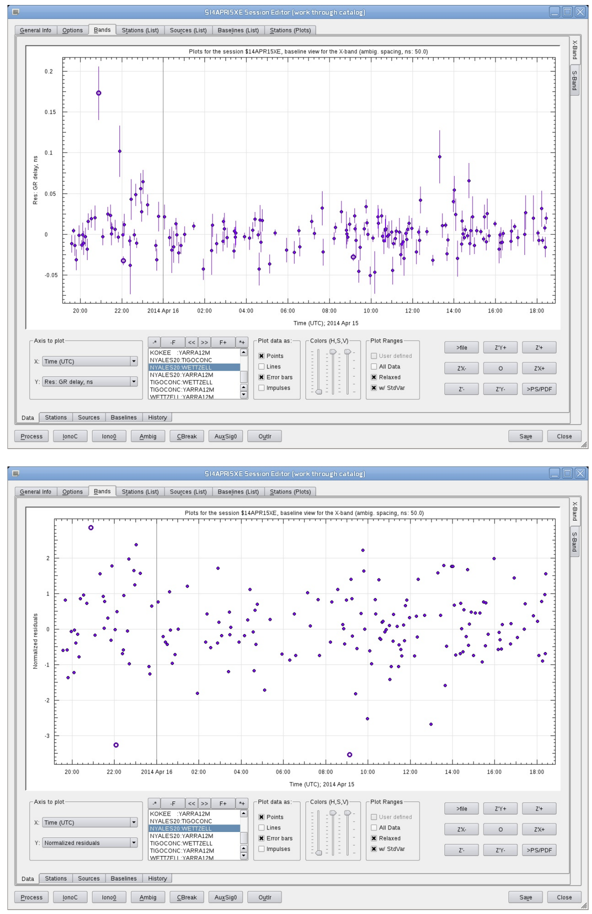

Fig. 3 Group delay residuals (upper screenshot) and the corresponding normalized residuals (lower screenshot) of the baseline NYALES20–WETTZELL obtained for the session \$14APR15XE.

Typically, the threshold for normalized residuals is 3 or 5. In some circumstances this corresponds to 3σ or 5σ thresholds of the residuals. The normalized residuals can be evaluated either for the whole set of processed observations or on a baseline basis. The process of outlier exclusion is an iterative process. After excluding an outlier from a solution, a new solution and normalized residuals are calculated. Also, the user can include previously excluded observations if the absolute values of their normalized residuals become less than the specified threshold. This action is called restoration and is performed iteratively. Both actions, outlier elimination and restoration, are performed in conjunction with the reweighting operation.

5 Conclusions

The first public release of the software νSolve was made in February 2014 along with a new CALC/SOLVE release. At the time of writing it can be accessed at the URL:

ftp://gemini.gsfc.nasa.gov/pub/misc/slb/

After the release, we have received user comments and suggestions and the software now is in a permanent process of updates. This is why it was released in a directory separate from the CALC/SOLVE directory. As for next major public releases, these will be distributed in the general CALC/SOLVE source tree.

In addition to the source code, a user guide on νSolve has been created. Currently it is available as a separate PDF file in the same FTP directory.

In the next releases we will focus on the following issues: 1) optimization of data processing time; 2) improvement of the plotting system; 3) extension of the functionality; and 4) introduction of the elements of automatic data processing.

References

1. G. J, Biermann. *Factorization Methods for Discrete Sequential Estimation. V128, Mathematics in Science and Engineering Series*, Academic Press, 1977.
2. S. Bolotin, J. M. Gipson and D. MacMillan. *Development of a New VLBI Data Analysis Software*. In D. Behrend, and K.D. Baver, editors, *IVS 2010 General Meeting Proceedings*, NASA CP 2010-215864, NASA GSFC, Maryland, pages 197–201, 2010.
3. S. Bolotin, J. M. Gipson, D. Gordon and D. MacMillan. *Current Status of Development of New VLBI Data Analysis Software* In W. Alef, S. Bernhart and A. Nothnagel, editors, *Proc. of the 20th Meeting of the EVGA*, Bonn, Germany, pages 86–88, 2011.
4. S. Bolotin, K. Baver, J. Gipson, D. Gordon and D. MacMillan. *The First Release of νSolve.* In D. Behrend, and K.D. Baver, editors, *IVS 2012 General Meeting Proceedings*, NASA CP 2012-217504, NASA GSFC, Maryland, pages 222–226, 2012.
5. S. Bolotin, K. Baver, J. Gipson, D. Gordon and D. MacMillan. *Current Status of νSolve.* In N. Zubko, and M. Poutanen, editors, *Proc. of 21st Meeting of the European VLBI Group for Geodesy and Astronomy*, p.77–80, ISSN: 0355-1962, Kirkkonummi 2013.
6. J. M. Gipson. *The Report of IVS-WG4.* In D. Behrend, and K.D. Baver, editors, *IVS 2012 General Meeting Proceedings*, NASA CP 2012-217504, NASA GSFC, Maryland, pages 212–221, 2012.
7. L. Petrov. *Memo about outliers elimination* CALC/SOLVE documentation, available at http://lacerta.gsfc.nasa.gov/mk5/help/elim_02_hlp.ps.gz
8. R. K. Rew and G. P. Davis. *NetCDF: An Interface for Scientific Data Access* IEEE Computer Graphics and Applications, Vol. 10, No. 4, pp.76–82, 1990.