Automatic Processing of Intensives at the NASA GSFC VLBI Analysis Center

Sergei Bolotin, Karen Baver, Mario Bérubé

Abstract In the middle of 2023, the NASA GSFC VLBI Analysis Center started processing the Intensive (INT) sessions in automatic mode. In this paper we describe the software used for this purpose (ADAP/APS and νSolve), the data flow from the release of a database by a correlator to the submission of the results of dUT1 estimation, and the experience of using the automatic processing of INT sessions at NASA GSFC.

Keywords VLBI, space geodesy, data analysis, dUT1, automation

1 Introduction

Very Long Baseline Interferometry (VLBI) Intensive (INT) sessions are conducted to determine the change in Earth rotation, which is measured as a correction to Universal Time (UT1). This correction varies unpredictably over time. UT1 is used in precise navigation, particularly GNSS. Also, estimating dUT1, the difference between Universal Time and Coordinated Universal Time, from VLBI observations is crucial for precise timekeeping, synchronization of global navigation systems, satellite operations, and astronomical observations. Finally, VLBI provides highly accurate measurements of celestial radio sources' coordinates, allowing for precise determination of the Earth's rotation parameters, including dUT1. These estimations help correct for irregularities in the Earth's rotation, ensuring the accuracy of timekeeping systems essential for various scientific, technological, and navigational applications. In all three cases, rapid turnaround is very important. The decrease of time elapsed from observations to obtained results can be shortened with the automatic data processing of new INT sessions at the analysis stage.

The importance of automatic data processing and low latency in VLBI observations for dUT1 estimation cannot be overstated. Automatic data processing enables rapid analysis of vast amounts of data collected from multiple observatories worldwide. This swift processing is essential for timely updates to timekeeping systems, ensuring accurate synchronization of global navigation networks, satellite operations, and scientific research. Automatic data processing leads to low latency, which guarantees that dUT1 estimations are available promptly, allowing for immediate adjustments to navigation systems and precise coordination of time-sensitive activities. Together, automatic data processing and low latency facilitate real-time decision-making and enable seamless operation of critical systems reliant on precise timekeeping.

In this paper we discuss automated data processing of INT sessions at the NASA Goddard Space Flight Center (GSFC) VLBI Analysis Center. Section 2 describes the data flow and ADAP/APS and νSolve software that are used for automated processing. In Section 3 we outline the experience of approximately the first nine months of our automated routine processing of INT sessions. Section 4 concludes that the automation cuts the primary latency (the interval of time elapsed since the last observation of a session until the results were obtained) for the IVS-INT-1 network approximately in half.

NVI, Inc./NASA Goddard Space Flight Center

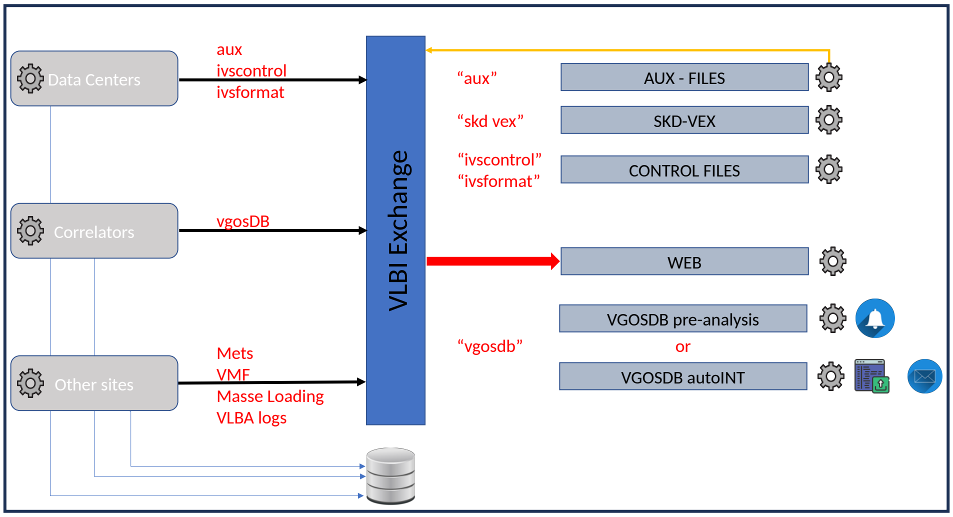

Fig. 1 Automated Data Acquisition and Processing (ADAP) system data flow.

2 Automation of Data Processing

2.1 ADAP/APS Software

The automatic processing of Intensives is part of the Automated Data Acquisition and Processing (ADAP) system, developed to support many GSFC VLBI activities. The ADAP system was built around a relational database management system MariaDB [4] for rapid access and validation of information and around an open source message broker software RabbitMQ [3] for controlling data flow between processes.

The system monitors IVS Data Centers, correlator ftp or http sites, and other specific servers for new data files required in processing VLBI sessions. The web crawlers, or scanners, are monitoring at regular intervals more than ten servers. When new information is detected, the scanner informs the message broker that dispatches the information via the 'VLBI Exchange' to the appropriate process using 'key words' based on file types. Each file type has one or many specific processes to validate and process it. Using many small applications facilitates maintenance and detection of problems. The same system maintains the IVS Sessions web pages with no duplication. Figure 1 shows part of the ADAP data flow related to analysis. Most of the processes ensure that all files required for processing are available and valid.

The 'VGOSDB autoINT' process is dedicated to automatic processing of Intensive sessions using νSolve for preliminary analysis of a database and the Automated Post Solve (APS) application for invoking the SOLVE software for creating solutions and submitting products and analysis reports as shown in Figure 2.

2.2 νSolve Software

To develop an application for automated analysis of an INT session, we used the script mode of the VLBI data processing software νSolve [1]. In script mode, νSolve reads commands from a script file and executes them. The script is a program written in the ECMAScript (standardized Java script [5]) computer language. The script engine is implemented in the Qt library.

We developed a script called autoINT.js that reads a database with a new INT session, processes it, and creates a new version of the database [2]. The script performs preliminary analysis, resolves ambiguities in each band, evaluates the ionosphere corrections, reweights observations, and handles outliers.

To control the quality of the performed analysis the script checks various criteria for validity of the solution: WRMS of the residuals, magnitude and standard deviation of the estimated dUT1, length of the session (an interval between the first and the last observation in

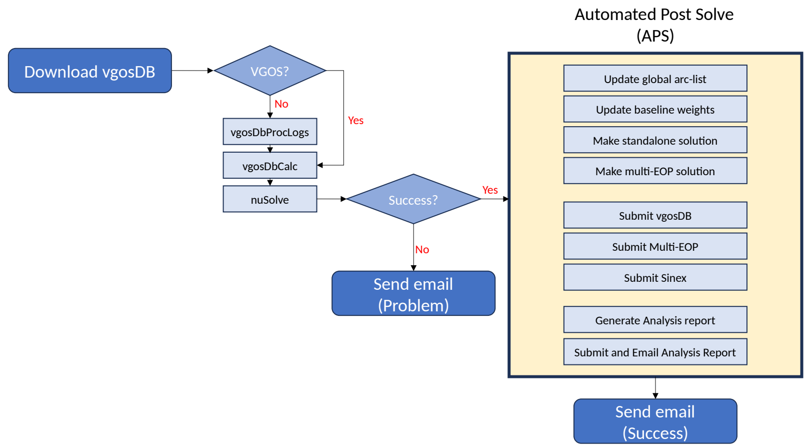

Fig. 2 Data flow of the automated processing of INT sessions at the GSFC VLBI Analysis Center.

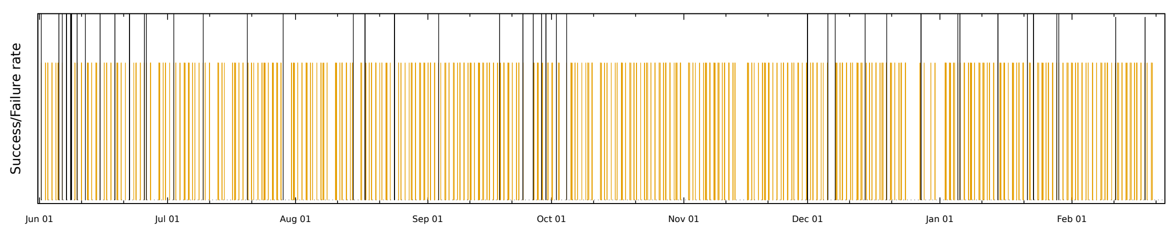

Fig. 3 Success/failure rates of the automatic processing of INT sessions starting in June 2023. Successful solutions are shown as orange short bars, and the sessions that required manual processing are shown as black bars.

the solution), number of observations rejected as outliers, and so on. If one of the parameters exceeds the threshold specified in the script, the script stops processing such a session, and it is then analyzed by an operator. In the case of success, the script saves the editing information as a new version of the database. This version of the database is usable for the batch mode processing of VLBI sessions.

The use of the autoINT.js script is integrated into the ADAP/APS software to process INT sessions in automated mode.

3 Use of the Automatic Processing at the GSFC VLBI Analysis Center

In April–May 2023, the GSFC VLBI Analysis Center undertook intense testing of the autoINT.js script. From the beginning of June 2023, routine processing of INT sessions has been transferred to automatic mode.

The interval of using the automated processing of INT sessions reported in this paper is from June 1, 2023 to February 20, 2024. During this period, 586 INT sessions were processed successfully by the autoINT.js script, and manual processing for 43 INT sessions was required.

Figure 3 shows the success/failure rates for the period since the beginning of the use of automated processing of INT sessions. The successful cases of the script execution are plotted with shorter bars to make the failure events more visible.

After the first four months of using the autoINT.js script, in October, 2023 it was revised. The sessions that the script was not able to process were analyzed, and adjustments and modifications to the script were made.

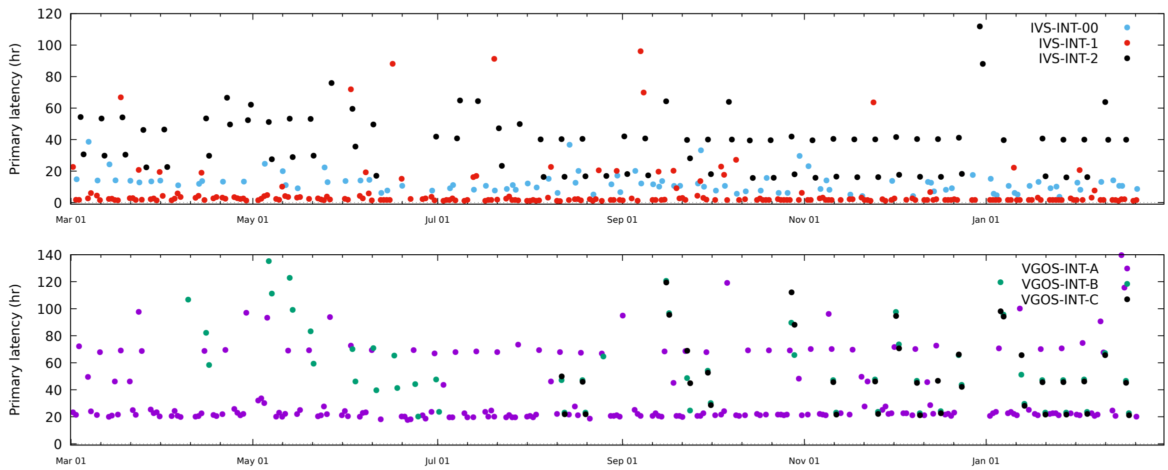

Fig. 4 Primary latency: interval of time from the last observation until the results were obtained. Top: selected S/X networks (IVS-INT-1, IVS-INT-2, and IVS-INT-00); bottom: selected VGOS networks (VGOS-INT-A, VGOS-INT-B, and VGOS-INT-C). Several points (sessions) of VGOS INT networks are not shown.

To evaluate the efficiency of the script, we calculated two types of latency times: primary and secondary.

We measure the primary latency as an interval of time elapsed since the last observation of a session until the results were obtained. This interval of time includes all processes that affect the VLBI observations: data transfer from stations to a correlator, processing the data at the correlator, transfer of a resulting database to one of the VLBI Data Centers, and analysis of the session by the GSFC VLBI Analysis Center. The primary latency characterizes overall performance of all the involved agencies. However, it does not include the interval of time that is necessary to transfer results of analysis to the Crustal Dynamics Data Information System (CDDIS) Data Center server, validation of the submitted files and copying of them into a publicly accessible place.

Figure 4 shows the calculated primary latency times for most of the major S/X (top plot) and VGOS (bottom plot) networks of INT sessions. The first sessions of each of the VGOS networks were processed in a testing mode and became available for analysis after one to two months after the observations. We did not plot these sessions because they change the scale of the plot.

For the INT networks that conducted enough sessions we evaluated the median values of the primary latency; these values illustrate typical latency of the networks. Tables 1–2 display the median latencies for the selected S/X and VGOS networks. The tables show

Table 1 Median values of the primary latency for the selected S/X INT networks before and after the use of automated processing started on June 1, 2023.

Period	IVS-INT-00	IVS-INT-1	IVS-INT-2	IVS-INT-S
pre 2023/6/1	14^h5^m	2^h54^m	$2^d4^h19^m$	$2^d8^h38^m$
from 2023/6/1	10^h5^m	1^h37^m	$1^d15^h54^m$	$1^d6^h16^m$

Table 2 Median values of the primary latency for the selected VGOS INT networks before and after the use of automated processing started on June 1, 2023.

Period	VGOS-INT-A	VGOS-INT-B	VGOS-INT-C
pre 2023/6/1	23^h38^m	$5^d2^h52^m$	$39^d0^h26^m$
from 2023/6/1	22^h15^m	$1^d23^h7^m$	$2^d4^h36^m$

the latencies for two periods: (a) from January 2023 to 31 May 2023 (before the automated processing of INT sessions was started) and (b) from 1 June 2023 through 20 February 2024. As one can see, the primary latency for all networks is better for the period of automated session processing. For example, the primary latency of the IVS-INT-1 sessions is shortened by almost one half, approximately from three hours to 1.5 hours. However, the reason for the drastic change in the primary latency for most of the VGOS INT sessions is not the automation of the analysis, but the fact that these networks were in the development stage in earlier observations and later became routine observations.

An interval of time from when the database was downloaded to the servers of GSFC until the results

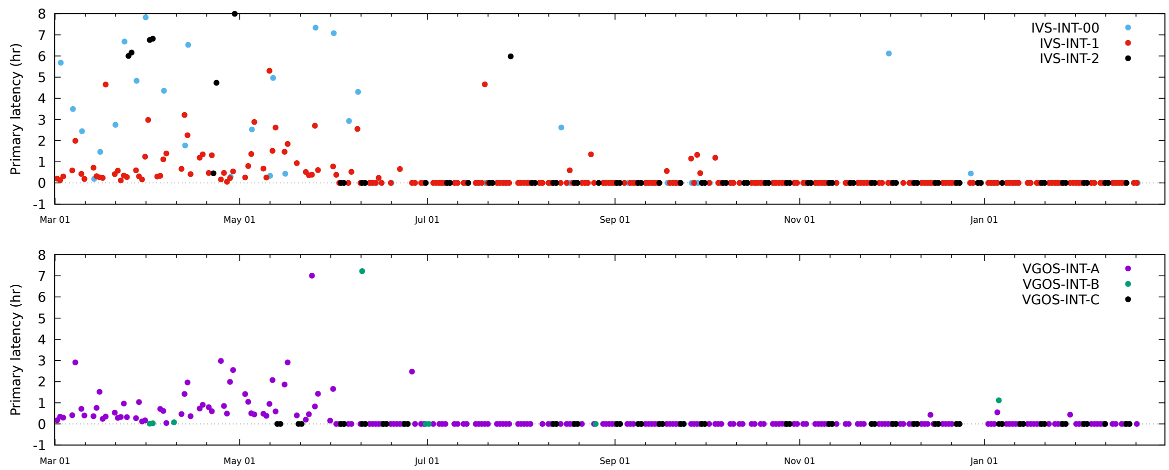

Fig. 5 Secondary latency: interval of time from when a new database was downloaded until the results were obtained. Top: selected S/X networks (IVS-INT-1, IVS-INT-2, and IVS-INT-00); bottom: selected VGOS networks (VGOS-INT-A, VGOS-INT-B, and VGOS-INT-C). Several points (sessions) on each of the plots are not shown.

were obtained we called a secondary latency. This value depends only on the data analysis process at the GSFC VLBI Analysis Center. It can vary from a few seconds in the case of successful automatic analysis to several hours or even days if the script was not able to process a session and involving manual data analysis was necessary.

Figure 5 shows the secondary latency for the S/X (top plot) and VGOS (bottom plot) networks of INT sessions. Some sessions are released on weekends. In the period before the beginning of automatic processing, these points on the plots have large latencies that distorts the scale and make other points indistinguishable; therefore we did not plot them.

Tables 3–4 display the median secondary latencies for the selected S/X and VGOS networks. Both the plots and the tables show significant improvements of the secondary latency for all INT sessions; they are changed from values on the order of hours or minutes to values on the order of seconds. It is especially noticeable for the IVS-INT-2, VGOS-INT-B, and VGOS-INT-C sessions, which are correlated at GSI, Japan (the facility's time zone is offset with respect to GSFC's time zone by almost 12 hours).

Table 3 Median values of the secondary latency of the selected S/X INT networks before and after the use of automated processing started on June 1, 2023.

Period	IVS-INT-00	IVS-INT-1	IVS-INT-2	IVS-INT-S
pre 2023/6/1	2^h32^m	28^m	8^h35^m	2^h19^m
from 2023/6/1	2^s	2^s	2^s	2^s

Table 4 Median values of the secondary latency of the selected VGOS INT networks before and after the use of automated processing started on June 1, 2023.

Period	VGOS-INT-A	VGOS-INT-B	VGOS-INT-C
pre 2023/6/1	29^m	14^h34^m	12^h45^m
from 2023/6/1	1^s	2^s	2^s

4 Conclusions

The software developed at the GSFC VLBI Analysis Center for automated processing of INT sessions now routinely analyzes the VLBI observations, estimates dUT1, and submits the results to the CDDIS server for future analysis and Earth rotation prediction.

The automated analysis of the INT sessions reduces the overall interval of time from the last observation of a session till the results become available. For the IVS-INT-1 network the primary latency shortened by almost half, approximately from three hours to 1.5 hours. However, for most of the INT networks the contribution of data analysis in the primary latency is not significant.

The secondary latency for all the INT networks dropped from hours to a couple of seconds. This is most noticeable when a session is released during non-business hours at GSFC.

In addition to reducing the latency of INT sessions, the use of automated processing has the following benefits: it relieves the personnel from the routine work, it eliminates user errors and it performs analysis of all the sessions in a uniform way.

The script autoINT.js is a part of the νSolve distribution. It is available starting with the nusolve-0.7.3 version of the distribution, and it is being improved with each new version. The distribution is available at the SourceForge site:

https://sourceforge.net/projects/nusolve

References

1. S. Bolotin, K. Baver, J.M. Gipson, D. Gordon and D. MacMillan. The VLBI data analysis software νSolve: development progress and plans for the future. In 8th IVS General Meeting Proc., ISBN 978-7-03-042974-2, Science Press, Beijing, China, pp. 253–257. 2014.
2. S. Bolotin, K. Baver, M. Bérubé and J. Gipson. Automatic Processing of INT Sessions with nuSolve. In 12th IVS General Meeting Proc., NASA/CP–20220018789, pp. 159–163. 2023.
3. L. Johansson and D. Dossot, RabbitMQ Essentials: Build distributed and scalable applications with message queuing using RabbitMQ, 2nd Edition, Packt Publishing, ISBN: 978-1789131666, 2020.
4. P. Mavro and F. Descamps, MariaDB High Performance, Packt Publishing, ISBN: 978-1783981601, 2014.
5. S. Stefanov, JavaScript Patterns, O'Reilly Media, Inc., ISBN 0596806752, 1st edition, 2010.