A. Müskens et al.: The Mark IV Correlator - Faster, Better, Optimal??, IVS 2002 General Meeting Proceedings, p.102–106
http://ivscc.gsfc.nasa.gov/publications/gm2002/mueskens

The Mark IV Correlator - Faster, Better, Optimal??

A. Müskens, I. Rottmann, H. Rottmann, A. Höfer, I. Benndorf

Geodetic Institute of the University of Bonn

Contact author: A. Müskens, e-mail: geodesy@mpifr-bonn.mpg.de

Abstract

We present a detailed efficiency analysis of the various processing stages of an experiment, from the tape shipment phase up to the final release of the correlated datasets. We summarize and evaluate the statistics accumulated for all geodetic experiments processed at the Bonn Correlator during 2001. These statistical data are used to identify "bottlenecks" and weak spots in the processing chain. We propose possible ways of further improving the efficiency of the Bonn correlator in order to match the IVS goals of fast and routine processing of geodetic experiments.

1. Introduction

The Bonn correlator is operated jointly by the BKG and the MPIfR and is available for processing of geodetic experiments for 50% of the total available time. In 2001 around 50 geodetic experiments of type IRIS-S, CORE-3, CORE-OHIG , CORE-C, and EUROPE were correlated in Bonn.

Starting in 2001 we have begun to statistically analyze all experiments in terms of how much time was spent in each of the various processing stages. Our goal is to identify the "bottlenecks" in the correlation process and propose ways of minimizing the time needed from observation to final data release.

2. Processing at the Bonn Correlator

The processing chain can be divided into the following stages:

- *shipment* - time between shipping the tape from the station and arrival at the Bonn correlator (see Section 3).
- *waiting queue* - time spent on checking the tapes into the local tape library, finding clock delays and waiting for free correlator time.
- *correlation* - time spent on processing, including overhead for tape positioning, synchronization, SU wait states etc.
- *post-correlation* - time spent on fourfitting the data, analysis, preparation of recorrelation files and waiting for free correlator time.
- *recorrelation* - time spent on reprocessing (including overhead).
- *database* - time spent on fourfitting the recorrelated data, analysis, refourfitting, creating a correlator report and submitting the data to the Data Center.

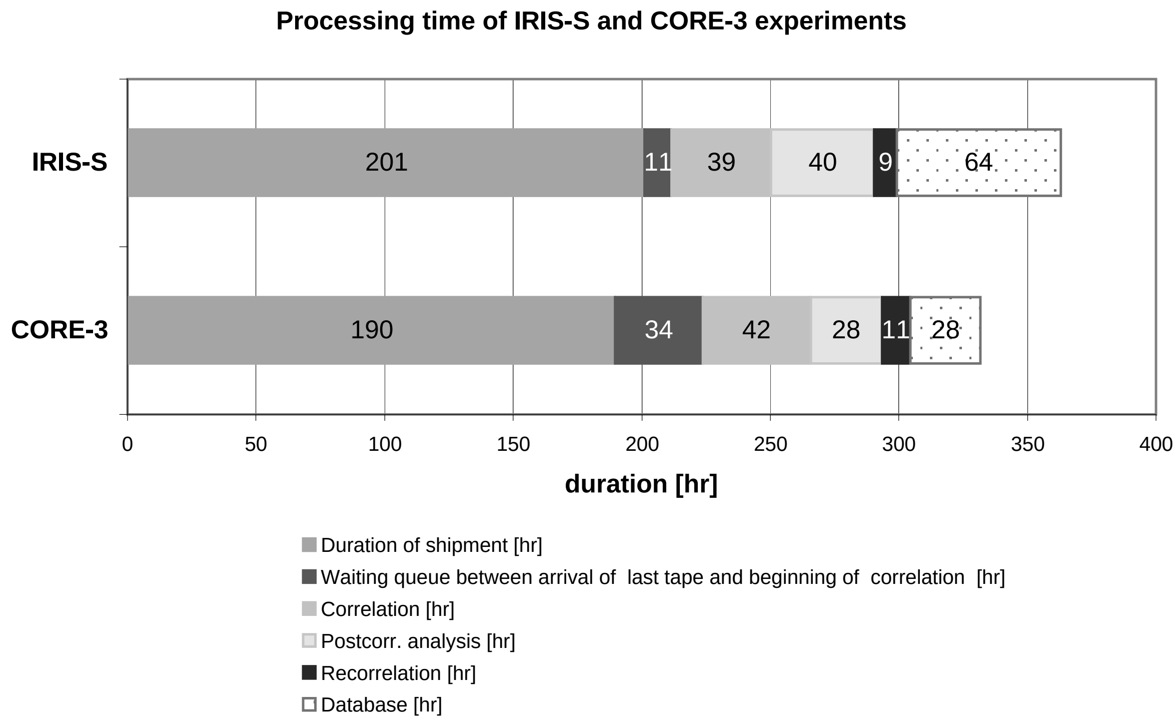


Figure 1. Time spent in the various stages of the processing chain for IRIS-S (top) and CORE-3 (bottom) experiments. Note that the average time for the database release was calculated excluding the C3007 to C3009 experiments. These experiments could not be released after completion due to a delay for a necessary software patch.

The results of the statistical analysis for CORE-3 and IRIS-S experiments are presented in Fig.1. As can be seen, the largest amount of time is spent on tape shipment ($\approx$ 200 hrs) which accounts for $\approx$ 50% of the total processing time. We will discuss the tape shipment in greater detail in Section 3 . Other time-consuming stages are the correlation and recorrelation. Combined, these steps require $\approx$ 50 hrs which is about a factor of 2 larger than the total duration of the observation (24 hrs). During correlation and recorrelation a considerable amount of time is spent on head positioning, tape changing, synchronization, SU wait states, etc.. These overhead tasks can amount to around 30% of the total correlation/recorrelation duration. The time required to release the database strongly depends on the quality of the observation and correlation. Before the database is released, data problems which have arisen have to be analyzed and solved (if possible). In the case of IRIS-S experiments the post-correlation analysis has resulted in release times of up to 60 hours.

3. Tape Shipment

The duration of tape shipment has been calculated separately for each experiment type and station. Note that due to the low number of CORE-C, CORE-OHIG and EUROPE experiments in 2001, we have excluded them from further analysis at this point. These experiment types will be statistically analyzed on a multi-year basis once more data becomes available. Remotely placed network stations like O'Higgins and Syowa naturally take a long time for tape shipment and consequently were not included in the analysis. For the calculation we have taken the day after the observations were finished as the first day of shipment. The day of arrival of the tapes at the Bonn Correlator was taken to be the last day of shipment. Furthermore, we have excluded Saturdays, Sundays and holidays, because usually no pick-up or delivery of the tapes is possible on these days. In the following we will refer to these as *shipment days*.

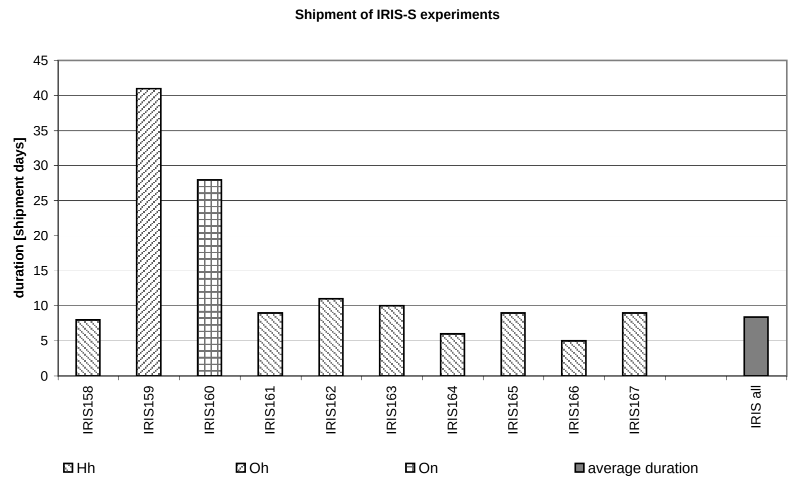
The results for IRIS-S and CORE-3 experiments can be found in Fig. 2. In the case of IRIS-S we have found that tapes from Hobart (Hh) take the longest (7 shipment days) to get to us. For CORE-3 setups including Medicina (Mc) and Onsala (On) these stations typically take the longest (9 and 6 shipment days respectively) to deliver the tapes to Bonn. CORE-3 setups including Ny Alesund (Ny) instead of Mc and On, result in an average shipment delay of 10 shipment days for Ny tapes.

4. Recorrelation - Is It Necessary?

After the initial correlation the fraction of "good" data is typically 80% - 90%. The most common data problems that require recorrelation of a particular scan are:

- the scan duration is too short for unknown reasons
- the scan was not correlated at all
- the scan has a quality code FQ = 0 for unknown reasons
- the scan has quality codes FQ = 1-5, A-H

Looking at the correlation phases outlined in Section 2 it is obvious that recorrelation can significantly decrease the efficiency of data processing. The data have to be inspected in detail prior to recorrelation. Setup files have to be produced and depending on the correlator load the waiting time until recorrelation can be long. Also the involved processing overhead due to tape changing, head positioning etc. is immense for recorrelation as compared to the initial correlation. If recorrelation has to be performed it is therefore desirable to reduce the number of recorrelated scans to a minimum. In order to achieve this goal we have developed a post-correlation analysis tool (*fifi*) that aids the analysts in attributing a particular data problem either to the observing station or the correlation process. By excluding bad station data from the recorrelation a considerable amount of processing time can be saved. On average about half of all "bad" scans are scheduled for recorrelation (around $10-20\%$ of the total number of scans). From these about 60% result in good data after reprocessing (see Fig.3 right). We can see that the fraction of bad scans that can be improved due to recorrelation is of the order of a few percent. Taking into account that the time needed for recorrelation is about 25% of the total correlation time (see Fig.3 left), we can conclude that recorrelation is highly inefficient. The processing factor (PF = processing time divided by observing time) of the recorrelation is about a factor of 3 higher than that of the correlation (see Fig. 4). We therefore seriously question the need for recorrelation.



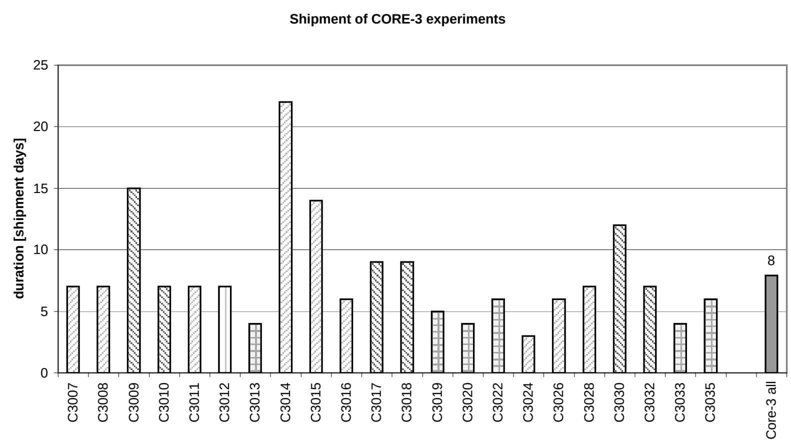


Figure 2. Shipment time for IRIS-S (top) and CORE-3 (bottom) experiments. Each bar corresponds to an individual station whith the longest deliver duration. The solid bar at the far right represents the average over all stations.

5. Conclusions

Based on the detailed analysis of the processing stages of all IRIS-S and CORE-3 experiments correlated in the year 2001 at the Bonn Correlator the following conclusions can be made:

- The tape shipment requires by far the largest amount of time in the processing chain. Network stations should commit to shipping tapes more efficiently, in order to reduce the delays to a minimum. This is very important and we strongly recommended that all network station strive for a speedy forwarding of recorded media.
- Recorrelation is inefficient. The amount of good data gained by recorrelation is small compared to the considerable amount of time spent on the recorrelation. We therefore seriously question whether recorrelation should be performed routinely in the future. Correlators should decide on the basis of the actual experiment whether recorrelation is reasonable or

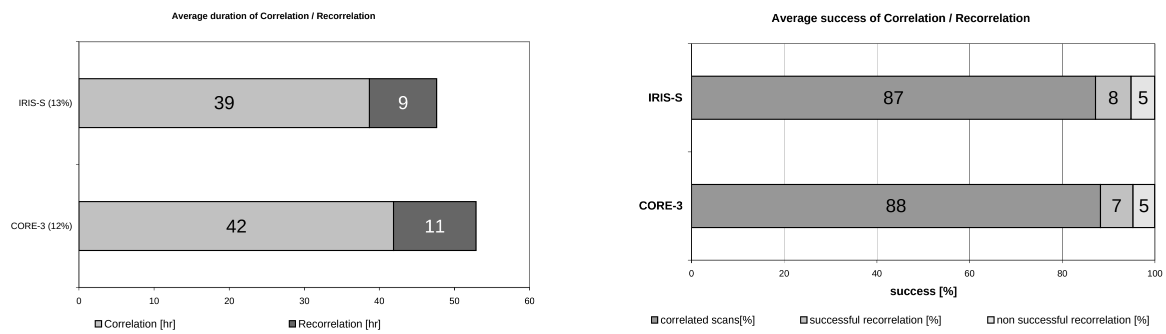


Figure 3. **left:** Total time [hrs] spent on the correlation and recorrelation of IRIS-S and CORE-3 experiments (the percentage corresponds to the fraction of recorrelated scans compared to the total number of scans); **right:** Amount of the good and bad data after correlation and recorrelation (as percentage of the total number of scans);

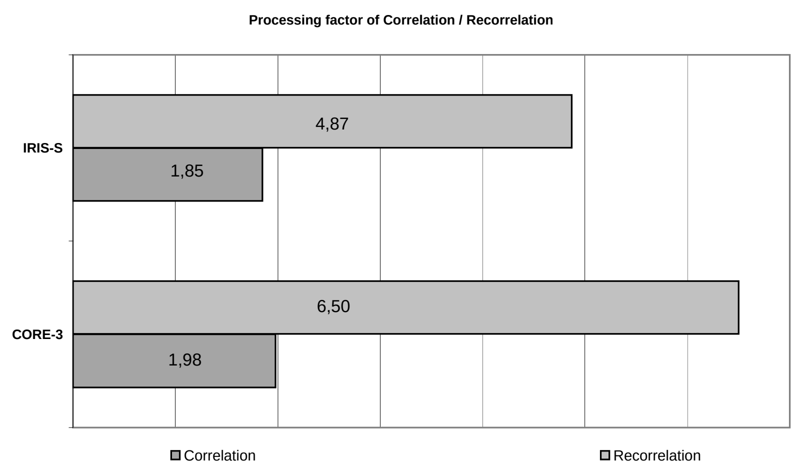


Figure 4. A comparison of processing factors of the correlation and the recorrelation. As can be seen the recorrelation results in processing factor that are worse by a factor of around 3. *Note: recorrelation PF in relation to the initial processing*

essential for further analysis.

- We hope to further speed-up the post-correlation data analysis by developing and improving software tools that aid the analyst in identifying and fixing data problems.
- The IVS recommendations for their products are clear and well constructed. We have to support these goals by challenging our (possibly well-established) procedures and rules.
- Bonn and possibly also other IVS Correlators should routinely collect and analyze efficiency information in the future. We propose to present similar analyzes at the next IVS meeting, to discuss our conclusions, and to possibly make recommendations in order to better meet the requirements of a rapid VLBI service.