SCIENTIFIC ANALYSIS OF DATA FROM A SCANNING SATELLITE [1]

C. Fabricius and E. Høg

Copenhagen University Observatory
Øster Voldgade 3, DK-1350 Copenhagen K, Denmark
E–mail: cf@astro.ku.dk, erik@astro.ku.dk

ABSTRACT

A scheme of astrometric and photometric data reduction for the GAIA mission is proposed. The reduction by several iterations of a small subset of all observations, of only one million *primary monitor stars* is sufficient to establish the final astrometric and photometric reference systems. Observations from the incoherent mode are reduced first. This solution is then used to set up initial values for the interferometric solution, which will give the definitive along scan attitude. A final reduction of the monitor stars in the incoherent mode gives the definitive z-axis attitude and the definitive calibrations. Subsequently, the 50 million programme stars are tied to this system in a single processing, without iteration of calibration and attitude parameters. – The first stages of astrometric reduction are similar to Tycho data analysis deriving the image location and signal amplitude from the raw observations. The resulting *raw transits* are used as input to the combined attitude reconstruction and great circle reduction and to the sphere solution which are similar to the Hipparcos data analysis. The final stage of tying all programme stars to the reference systems is similar to Tycho astrometry and photometry. We conclude that the required data reduction is a manageable, albeit very large, task.

Keywords: space astrometry; GAIA; data reduction.

1. INTRODUCTION

The data reduction consortia for the Hipparcos-Tycho project have demonstrated that the reduction of such amounts of data as from that project is a very large but manageable task. The proposed GAIA mission (Lindegren & Perryman 1994, hereafter LP) presents a data reduction task several times larger than that of the Hipparcos-Tycho project. We will here try to demonstrate that although the data reduction will be much more demanding than what has been seen up to now, it may in fact be carried out provided a carefully planned startegy is followed.

The GAIA concept contains an interferometric part of high precision and an incoherent part with lower precision but a fainter limiting magnitude and a better light economy. We will base our discussion on this proposal and assume that a programme of 50 million stars is observed by both parts although the incoherent part could be extended significantly due to the better limiting magnitude.

LP leaves many options open for the detailed detector design and we will therefore limit ourselves to rather general considerations specially when referring to the interferometric part. The data reduction for the two parts will follow much the same lines, but they interact in several ways. A first solution for the incoherent part is needed to remove ambiguities in the interferometric solution. It is also needed in order to determine the attitude of the z-axis, i.e. the satellite body axis parallel to the spin axis. The reduction of the interferometric observations will, on the other hand, provide the definitive attitude values along scan to be used in the final reduction of all observations.

It is at the moment unclear what auxiliary data will be needed and in what amount. We will therefore not discuss the metrology data needed in the reduction.

The scientific results from the GAIA data reduction for each star are the five astrometric parameters, possibly perspective acceleration and orbital parameters and the magnitude in several colour passbands limited by the spectral response of telescope and CCD. The intensity of the sky background is obtained in the same bands. Photometric mean values for all stars will be made available as well as photometric results for each crossing of a CCD for selected stars.

The present paper is an update of an earlier paper (Høg 1993) discussing data reduction for the ROEMER proposal.

2. OBSERVATIONS WITH GAIA

The raw observational data from the satellite consist of a sequence of photon counts for each star at each crossing of a CCD. The star is identified by its number (ID) in the GAIA Input Catalogue (IC) of 50 million stars which may be expected to have an accuracy of position of $0\overset{''}{.}25$ rms at the epoch of observation. The samples are essentially equidistant in time and the on-board time of each sample should be available with a resolution better than 0.1 microsecond.

[1] A contribution to *Future Possibilities for Astrometry in Space*, a workshop at RGO, Cambridge, England, 19-21 June 1995

The raw observations are analysed on the ground using a priori (pre-launch) knowledge of the geometry of the telescope and CCDs and of the spectral passbands. The Real Time Attitude (RTA) used on board is included in the raw observations. This is sufficient to derive the geometric and photometric calibration parameters as functions of time and thus the final scientific results.

2.1. Coherent and Incoherent Modes

Detector systems for the *coherent*, interferometric observations with GAIA have been described by Lindegren & Perryman (1994), and detailed designs have been given by Høg (1995) for which data rates are given in the following Section. The focal plane of each of the three telescopes contains 28 CCD chips for coherent imaging, each $540''$ square.

For *incoherent* imaging the focal plane of each of the three telescopes contains 20 CCD chips, each $680''$ high. With a spin rate of $2'$/s these 60 CCD chips will scan 1360 $\square'$/s giving each star an average of 1445 observations (or CCD chip crossings) during 5 years. For a programme of 50 million stars we get an average of 458 such chip crossings each second. The CCD chips have a wide CCD region with small pixels ($0.''1 \times 4''$) for observations of all stars but the brightest. They have also a narrow CCD region with larger pixels ($0.''1 \times 12''$) allowing brighter stars to be observed, but less suited for very faint stars or stars in very dense fields. For simplicity, we will assume that all stars are observed on both parts of the CCD chip.

Predictions of chip crossings must be available on-board. They are based on the Input Catalogue and the Real Time Attitude. A raw observation contains data from two neighbouring strips of CCD pixels, one above the other, each $3''$ (32 pixels) long. Two strips are in general needed because many stars cross from one strip to the next during an observation and a length of $3''$ allows for IC errors, image size and RTA errors.

2.2. Data Rates

For the interferometric (*coherent*) mode Høg (1995) has estimated the data rate to **360 kbit/s** for a specific detector design and preprocessing strategy.

For the *incoherent* mode only one design has been proposed. Based on the Input Catalogue and the RTA we can safely assume all relevant information for a crossing to lie within a specific sequence of 32 pixels ($\sim 3''$). A star will often affect two rows of pixels, so when including both CCD regions of the chip, $4 \times 32 = 128$ pixels must be transmitted to ground for each CCD chip crossing. With an average of 458 crossings/s, with 12 bits/pixel and with 160 bits for a star header (cf Høg 1995), the data rate becomes 780 kbit/s. This very high rate can be reduced significantly at the cost of some on-board preprocessing. A simple detection algorithm could identify which 10 pixel sequence out of the 32 is relevant. Furthermore the narrow CCD region has rather high pixels and in general only one pixel sequence will be of interest from that region. All this brings us down to 30 pixels per chip crossing, resulting in an average data rate of **240 kbit/s** for the incoherent mode, which is probably acceptable.

Table 1: Estimate of the amount of data for the incoherent mode from a 5 year GAIA mission, compared with 2.5 years of Tycho mission (Høg et al. 1992). It is assumed that an average of 30 pixel values are transmitted to the ground for each CCD chip crossing. (Gb = Gbyte.)

	Tycho		GAIA Incoh.		
			Monitor		Progr
	1 mio		2 mio		48 mio
	1 obs		1 obs		
	byte	Gb	byte	Gb	Gb
RO Raw Obs.	-	100	60	175	4200
BG Background					20
RT Raw Transits	64	80	64	185	4400
TC Transit Cat.	68	8	64	185	4400
STC Sorted TC	68	8	44	125	3100
APOC Output Cat	-	0.3		1	15
Data archive				670	16100

3. THE DATA REDUCTION

In the discussion of the data reduction we will be referring to the following data sets:

IC	Input Catalogue of 50 million stars
ICM	Input Catalogue of 2 million Monitor stars
RO	Raw Observations
BG	Background
RT	Raw Transits
TC	Transit Catalogue of individual calibrated transits
STC	Sorted Transit Catalogue
APOC	Astrometric and Photometric Output Catalogue

The approximate sizes of these data sets have been estimated in Table 1 for the incoherent mode. For the interferometric mode the amounts of data will be similar.

The amount of data to be reduced will be several thousands of gigabytes. Carrying out a reduction scheme similar to the Hipparcos reductions seems hardly possible. It is, however possible to carry out all calibrations and data checking with only a small subset of all observations. The observations of 1–2 million *monitor stars* are sufficient, consisting of a few hundred gigabytes of raw data in 5 years. This is a few times the amount of raw observations from 2.5 years of the Hipparcos and Tycho missions and should be fairly easy to handle even with the computing facilities of today.

The monitor stars should consist of one million bright stars and one million fainter objects down to the limiting magnitude including photometric standards, quasars, galaxies, major planets and their satellites, minor planets, all known and suspected variable stars, known and suspected double stars, nearby stars, high velocity stars and other objects known to be a challenge.

The monitor stars are divided in two groups. The *Primary Monitor Stars* are used to establish all calibrations and to establish great circle and sphere solutions. The

Secondary Monitor Stars are used to test the implementation of this solution before starting the routine processing of the bulk of programme stars. The primary monitor stars are a subset of about one million monitor stars suited for defining the astrometric and photometric reference systems. They should consist of a uniformly distributed grid of single stars, mainly, but not exclusively, in the magnitude range of highest accuracy, i.e. $V < 11$. The stars should be single in the sense that other objects within a distance corresponding to the size of a GAIA interferometric subfield, are sufficiently faint not to produce significant disturbances. Makarov *et al.* (1995) discuss some aspects of the data reduction for the proposed ROEMER satellite. They show that with two fields of view, each $1^\circ \times 0.9^\circ$ and with a set of 760 000 stars on the sky, the abscissæ determined in the great circle solution will have random errors only slightly higher ($\sim 6\%$) than the theoretical minimum. Such a set of stars will also suffice for the sphere solution. For GAIA, with three large fields of view, the situation is even more favourable, and we may safely assume that a set of nearly 1 million stars will be sufficient. Therefore, these stars and perhaps 100 000 fainter monitor objects especially quasars define the subset of *Primary Monitor Stars* adequate for all iterations of calibrations, great circle reductions and sphere solutions.

In the following paragraphs we describe in general terms the flow of data in the various reduction steps. The flow diagrams show input to and output from the processes. Single or double arrows indicate single or repeated input:

Explanation:	$\rightarrow$	single input of values
	$\Rightarrow$	repeated input of iterated values

3.1. Production of Raw Observations

The first step is to separate the raw observations into three groups: Raw Observations of Primary Monitor Stars (ROPM), Raw Observations of Secondary Monitor Stars (RO2M) and Raw Observations of the Programme Stars (RO). This separation is based on flags in the input catalogue. It will facilitate the later processing, if the observations of programme stars are sorted according to the ID star number at least to some extent.

Telemetry from Satellite	$\rightarrow$	Separation of Raw Observations of Monitor Stars	$\rightarrow$	ROPM 90 Gb
			$\rightarrow$	RO2M 90 Gb
			$\rightarrow$	RO 4200 Gb

The on-board selection of pixels to be transmitted is based on the Real Time Attitude and the Input Catalogue. The raw observations are labeled with this preliminary identification. They include the RTA, time, scan velocities, identification of pixels and pixel values.

3.2. Background from all Raw Observations

The background countrate can normally be derived from the individual observation giving a *local background*. Mean values of such local backgrounds give an *area background* which replaces the local background in crowded fields.

RO Raw Obs.	$\rightarrow$	Background Determination	$\rightarrow$	BG 20 Gb

The production of background values can conveniently be combined with the separation of raw observations.

3.3. Iterations of Primary Monitor Stars

Iterations are carried out first for the incoherent part. This gives the z-axis attitude and an improved input catalogue. The interferometric part may now be reduced without ambiguities, giving the definitive along scan attitude. Finally the incoherent part is reduced again using this final attitude. Geometric and photometric calibration parameters for each CCD are also determined in this process.

The first step is the revised prediction of each CCD crossing, the detection of a star and the estimation of the signal amplitude. This is done for one raw observation at a time. To the extent detection and estimation have been carried out on-board, no revised values will of course be produced.

Intermediate data as for instance the predicted chip crossings, resulting from the prediction process, might be saved on tape for later use in order to save a recalculation. But we have learnt from Tycho data analysis (Høg et al. 1992) that the maintenance of correct interfaces for such data is so laborious that a recalculation should generally be preferred, given the fast computers of the future.

ICPM Input Cat.	$\rightarrow$	Prediction, Detection and Estimation	$\rightarrow$	RTPM Raw Transits 90 Gb
Ephemerides	$\rightarrow$			
ROPM Raw Obs.	$\rightarrow$			
BG Background	$\rightarrow$			
Calibr. Par.	$\Rightarrow$			
Attitude	$\Rightarrow$			

These *Raw Transits* include: ID, predicted and estimated transit time, scan velocities, CCD, z-coord., position angle of CCD, mag in IC, amplitude, local and area background, quality estimates and flags.

In the second step, the raw transits from a few revolutions of the satellite are used to determine the attitude of the satellite and the abscissæ of the stars along a reference great circle. The attitude is an input to this process when definitive attitude values have been determined.

RTPM	$\Rightarrow$	Attitude and Great Circle Reduction	$\rightarrow$	Attitude
Calibr. Par.	$\Rightarrow$		$\rightarrow$	Abscissae
AOCPM	$\Rightarrow$		$\rightarrow$	POCPM Photometric Output Cat.
POCPM	$\Rightarrow$			
(Attitude)	$(\rightarrow)$			

The third step is the *Sphere Solution* producing the final astrometric results from the reference great circles and star abscissæ. It may well be possible to combine the great circle reduction and the sphere solution into a single *Global Solution*. Experiments by C. Petersen (1995, priv. comm.) using the Hipparcos observations look promising.

Attitude	⇒	Sphere Solution	→	AOCPM
Abscissae	⇒			Astrometric Output Cat

In the last step of the iteration, the geometric and photometric calibration parameters are improved before starting the next iteration.

AOCPM	⇒	Geometric and Photometric Calibration	→	Calibr. Par.
RTPM	⇒			
Attitude	⇒			

3.4. Final Reduction of the Primary Monitor Stars

ICPM	→	Final identification and sorting into regions	→	APOCPM 0.3 Gb
POCPM	→			Astr & Phot.
AOCPM	→			mean values
Ephemerides	→			
RTPM	→		→	TCPM 90 Gb
Calibr. Par.	→			Transit Cat. of
Attitude	→			indiv. values

TCPM	→	Sorting of Transits	→	STCPM
				Sorted TCPM, 65 Gb

3.5. Reduction of Raw Observations of other Stars

After complete iterations of the primary stars follows the final reduction of the remaining, secondary, monitor stars resulting in catalogues containing mean values for each star and the individual calibrated values for each star transit. The reduction of the one million secondary monitor stars will provide a thorough check on the correctness of data and software for the reduction of all 50 million programme stars since this will be done with the same software.

IC Input Cat.	→	Prediction of Chip Crossings & Detection and Estimation		
Ephemerides	→		→	RT
Calibr. Par.	→			Raw
Attitude	→			Transits
RO Raw Obs.	→			4400 Gb
BG Backgr.	→			

RT	→	Transit Identification	→	APOC 15 Gb
IC Input Cat.	→			Astr. & Phot.
Calibr. Par.	→			mean values
POCM	→		→	TC 4400 Gb
AOCM	→			Transit Cat.
				of indiv. values

TC	→	Sorting of transits	→	STC Sorted TC
				3100 Gb

This reduction scheme applies to the secondary monitor stars as well as the programme stars. The sorting may be restricted to a subset of stars.

The individual calibrated transits are written in the transit catalogue (TC) and the mean values are given in the astrometric and photometric output catalogue (APOC). The calibrated transits are sorted according to star number to allow a single step solution and to allow analysis of variable stars and of solar system objects. This sorting is essential for the monitoring stars and desirable for the programme stars although it may prove to require too large an effort. Alternatively, a part of the sorting could be performed during the production of raw observations on the ground, i.e. during the mission.

4. CONCLUDING REMARKS

Two data reduction consortia should be planned as for the Hipparcos main mission in order to ensure the reliability, especially of the complex astrometric processing. It would be sufficient if the two consortia would process and compare results for the two million monitor stars. It is fairly simple to connect the remaining 50 million programme stars to the firmly established astrometric and photometric systems of monitor stars. This needs therefore only be carried out at one place, but, alternatively, several institutes could share the work.

ACKNOWLEDGEMENTS

This work was supported by the Danish Space Board. One of the authors (EH) is much indebted to members of the Hipparcos Science Team for discussions on the data reductions for a Hipparcos-2 project and for the proposed ROEMER mission. The authors are grateful to Drs U. Bastian and A. Wicenec for discussions of a proposed second Tycho processing and to Dr V.V. Makarov for comments to a previous version of this paper.

REFERENCES

Høg E., Bastian U., Egret D., Grewing M., Halbwachs J.L., Wicenec A., Bässgen G., Bernacca P.L., Donati F., Kovalevsky J., van Leeuwen F., Lindegren L., Pedersen H., Perryman M.A.C., Petersen C., Scales D., Snijders M.A.J., Wesselius P.R., 1992, A&A, 258, 177

Høg E. 1993, Data reduction for a ROEMER mission, in: *E. Høg (ed.) Annex B to the ROEMER proposal*, p19

Høg E. 1995, Some designs of the GAIA detector system, in these Proceedings

Lindegren L., Perryman M.A.C. 1994, GAIA – Global Astrometric Interferometer for Astrophysics, in *Supplementary Information Submitted to the Horizon 2000+ Survey Committee*, 52pp.

Makarov V.V., Høg E., Lindegren L. 1995, Random errors of star abscissae in the ROEMER space astrometry project, *Experimental Astronomy* (in press)