Algorithms for GDAAS Phase II: Definition

GAIA Algorithm Community

Version 4.1 10th July, 2003

Abstract. In the second phase of the GAIA Data Access and Analysis Study (GDAAS-2) a set of representative algorithms will be implemented in the areas of astrometry, photometry, radial velocity, minor planets and classification. This document provides a concise description of proposed algorithms in terms of functionality, method, input/output data, and data access requirements. A proposed calibration model for Astro is described in an Appendix. — Version 4 of this document takes into account numerous comments by the GST and WG leaders on the previous version. The flow charts have been removed; instead, data access requirements are summarised for each algorithm.

Contents

0 Introduction 2

1 Astrometry 4
   1.1 Improved object matching (1-01-C) 4

2 Proposed Area 7
   2.1 Title 7
   2.2 Fundamental algorithms (1-03-C) 9
   2.3 Improved attitude modelling (1-04-C) 13
   2.4 Improved instrument modelling (1-05-C) 15
   2.5 Detailed geometrical calibration (1-06-C) 19
   2.6 Chromaticity calibration (1-07-C) 21
   2.7 Improved CCD calibration (1-08-C) 23
   2.8 Additional global parameters (1-09-S) 24
   2.9 Visual double star analysis (1-10-S) 26
   2.10 Astrometric Binary star analysis (1-11-S) 28

1This document was originally authored by L. Lindegren. It has been revised by several authors for each individual algorithm ever since the last version (Version 4) of the original document.
2.11 Astrometric planet detection (1-12-S) ........................................... 30
2.12 Astrometry quick-look analysis (1-13-Q) ...................................... 32
2.13 Photometric calibration (2-01-C) .................................................. 35
2.14 Imaging analysis (2-02-S) ............................................................. 39
2.15 Variability analysis (2-03-S) ........................................................ 41
2.16 Science alerts (2-04-Q) ............................................................... 43

3 Radial Velocity 44
3.1 Radial velocity source detection (3-01-C) ....................................... 44
3.2 Radial velocity wavelength calibration (3-02-C) ............................... 46
3.3 Radial velocity cross-correlation (3-03-S) ....................................... 48

4 Minor planets 53
4.1 Minor planets motion detection (4-01-C) ......................................... 53
4.2 Minor planets cross-identification (4-02-S) ..................................... 55
4.3 Minor planets orbital solution (4-03-S) .......................................... 56

5 Classification and Parametrization 58
5.1 Ingestion level classifier, ILC (5-01-C) ............................................. 60
5.2 Multiple source detector, MSD (5-02-C) ......................................... 62
5.3 Discrete source classifier, DSC (5-03-C) ......................................... 64
5.4 Single Star Parametrizer, SSP (5-04-S) .......................................... 66
5.5 Single quasar parametrizer, SQP (5-05-S) ....................................... 70

Appendix A: Overview of Astro calibration parameters 72
A.1 Proposed calibration units, temporal and spatial scales ..................... 72
A.2 Use of a standard (instrument) photometric system .......................... 74
A.3 Geometric and chromatic calibrations ........................................... 74
A.4 Response function (LSF/PSF) calibration ....................................... 77
A.5 Photometric calibration ............................................................... 79
0 Introduction

The first phase of the GAIA Data Access and Analysis Study (GDAAS) was completed in mid-2002. It demonstrated the general feasibility of the application of a database architecture to the organisation, archiving and analysis of GAIA data. The study included the implementation and testing of a few central algorithms deemed critical for the approach, viz. the ingestion of raw satellite data, object matching, and the global iterative solution.

GDAAS Phase II (GDAAS-2 for brevity) was started in July 2002 and is estimated to run for 2.5 years. In this time frame, it is expected that a considerable number of additional algorithms will be implemented within the database/data analysis system. A primary purpose is to arrive at a realistic assessment of the target system in terms of complexity, storage and processing requirements. However, it can also be viewed as an exercise in methodology, viz. how to go about defining, developing, configuring, integrating and testing user-supplied algorithms within a very complex and strictly controlled data processing environment.

The Statement of Work for GDAAS-2 (27 June 2002) contains a list of proposed algorithms expected to be integrated during this phase. The present document provides a concise definition of each of these algorithms by means of the following elements:

1. Function: a short description of what the algorithm is supposed to do
2. Method: a short description of how the algorithm works
3. Input data
4. Output data
5. Data access requirements: a description of READ and WRITE patterns for each of the main data types
6. Remarks
7. Provider: who is responsible for providing the algorithm
8. Schedule: delivery milestone(s)

The first four items define what the algorithm is supposed to do. In contrast, the item data access requirements provides a first-order assessment of how the algorithm will interact with the database system. This may take into account processes that are not part of the algorithm itself, as described. For example, the improved object matching uses simple lists of detections and sources as its main input and output. This is a simplification both for defining and testing the algorithm. However, when the object matching algorithm is run, other processes are used to create and interpret the lists, and it is the access requirements of these processes that are described under this item.

Requirements for reading and writing to the database are summarised in the form:

\[ \text{Type of data} \]
\[ \begin{align*}
\text{READ} & \quad \text{how many} / \text{access key(s)} / \text{further selection} \\
\text{WRITE} & \quad \text{how many} / \text{selection} / \text{update or insert}
\end{align*} \]

\[ ^2 \text{It is at present not entirely clear who is responsible for providing algorithms for these other processes.} \]
The following types of data are considered:

- observations (raw CCD samples; *not to be updated*)
- elementary data (centroid, flux and background for an observation)
- sources (i.e., astronomical objects: stars, quastars, minor planets, ...)
- attitude (quaternions as function of time)
- calibrations (geometric, photometric, chromaticity, LSF, PSF)
- global parameters (PPN parameter $\gamma$ and additional parameters)
- ephemerides (satellite orbit, major planets; *not to be updated*)
- minor planets orbital elements
- auxiliary data (sky history, spectrum templates, ...)

Relevant access keys are:

- time: observations, elementary data, attitude, calibrations, ephemerides
- source ID or position (HTM node): observations, elementary data, sources, minor planets orbital elements
- CCD number: observations, elementary data, calibrations
- links from other data types (TBD)
1 Astrometry

1.1 Improved object matching (1-01-C)

1.1.1 Function

To assign a source to a given detection (in any of the skymappers). If no matching source is found in the database, a new source is created.

1.1.2 Method

The matching is based on coincidence in celestial coordinates, using current attitude and calibration data. A matching source must be found within a pre-defined search radius of the detected position; if there are multiple sources within that radius, the matching source is the one having the minimum distance to the detected position. ‘Distance’ (and ‘radius’) may be defined by a non-isotropic metric, corresponding to the different coordinate precisions along and across scan.

1.1.3 Input data

- control parameter: search radius
- a list of detections with celestial coordinates and position errors
- a list of sources already in the database, within the scanned area: source ID, predicted coordinates with uncertainties
- an approximate center of the scanned area

Note: The above lists refer to a particular stretch of skymapper data to be processed, which may be defined by the instrument, range of CCDs, and time interval (beginning and end time).

1.1.4 Output data

- for each detected source, the matching source ID, or 0 if no match was found

1.1.5 Data access requirements

Elementary data:
READ batch / time and CCD / special: those not linked to source
WRITE batch / all read / links to sources updated
1.1 Improved object matching (1-01-C)

Sources:
- READ batch / position
- WRITE batch / subset of those read + new sources / update and insert

Attitude:
- READ batch / time
- WRITE none

Calibration:
- READ batch / time and CCD
- WRITE none

Global parameters:
- READ all
- WRITE none

Ephemerides:
- READ batch / time
- WRITE none

1.1.6 Remarks

1. The object matching (OM) algorithm is used for the (quasi-continuous) pre-processing of data from all skymappers (ASM, MBPSM, RVSM). It performs the (initial) cross-matching of all kinds of sources, including solar-system bodies. However, the algorithm could also be re-run e.g. after a provisional source catalogue and/or updated attitude and calibration data have been established in order to clean up the catalogue.

2. This version is based on the OM algorithm adopted for GDASS-I, which performs a recursive first neighbor search in a 2D plane as defined by the standard coordinates derived from the input coordinates (RA,DEC). The adoption of the “equidistant” projection in place of the usual gnomonic projection makes it possible to operate on a very large field of view (up to a few tens of degrees in size). In addition, the transformations of coordinates do not appear critical in term of the time requests of this task which appears dominated by the database operations, according to what we inferred from conversations with X.Luri. The main change of this new version is the adoption of a data module instead of a temporary file that is used to exchange the matching results among the various subroutines, and which also increases the speed of the OM code. The adoption of the detector coordinates in place of the current standard coordinates, as proposed by L.Lindegren (2003, GAIA–LL–044), may be implemented in future versions of the algorithm if necessary to improve the efficiency of the OM algorithm.

3. According to the implementation schema described by Lattanzi, Luri et al. (2001, GDAAS-TN-005, v1.2), the OM algorithm is based on the assumption that the input list of new observations does not include multiple detections of the same source. To this regard, recently, X. Luri remarked that the database accesses will be significantly improved if multiple scans (i.e. with multiple observations) could be processed by the OM code. This is problem of data processing optimization related to the database architecture which can
be managed by the database interface, although a change of the OM algorithm and a redefinition of this task can also be considered for the future releases.

4. Following comments by the Torino group (2003, GAIA–ML–018.1) the definition of an ‘optimal’ object matching algorithm (based on a probabilistic approach taking into account uncertainties in the input data as well as intensity, shape, etc.) is postponed to a later stage of the software development.

5. It is also assumed that predicted coordinates (and errors) of the objects in the Source list are updated to the observing epoch via the Fundamental algorithms or that the effects due to not well estimated parameters (eg. proper motion, gravitational deflection, etc.) are negligible. Basically, the object matching does not need to attain the full (sub-µas) precision of this calculation, although the use of a unique fundamental algorithm is strongly recommended unless it is found wholly impractical (cf. Sect. 2.2).

6. For new sources the approximate source parameters (position, magnitude) need to be computed from the elementary data, through an inversion of the calibration and attitude, and removing aberration.

7. The procedures described under items 5 and 6 above are not part of the matching algorithm proper.

1.1.7 Provider

Spagna A.

1.1.8 Schedule

Algorithm delivery: July 2003
2 Proposed Area

2.1 Title

2.1.1 Function

To perform the on-ground ASM data handling, in order to provide improved parameters such as position (especially transverse position), flux, shape and background, using colour information and calibrations (especially PSF) not available for the on-board ASM data handling.

2.1.2 Method

Fitting of background plus PSF to ASM samples, using current source and calibration data from the database. The fitting will be performed using maximum likelihood (see GDAAS-FA-01) or robust weighted least-squares (description to be written) algorithms. A choice will be made between these two methods according to the results of intensive simulations.

2.1.3 Input data

- control parameters: ASM number and time interval to process
- ASM samples in the selected interval
- source parameters for observed sources
- calibration data (PSF) for ASM

2.1.4 Output data

- improved pixel positions and fluxes of the sources
- shape parameters
- background estimates
- statistics (goodness of fit, covariances)

2.1.5 Data access requirements

Observations: READ batch / time and CCD / only ASM WRITE none
Elementary data: READ batch / same as observations WRITE batch / all read / update
Sources: READ batch / link from elementary data WRITE none
2.1 Title

Attitude: READ batch / time WRITE none

Calibration: READ batch / time and CCD / only ASM WRITE none

2.1.6 Remarks

1. ASM data have the unique advantage of giving information (sources, local background) in a 10x10 or 12x12 pixel area for each FOV separately; otherwise, such information is always mixed from both FOV in the Astro focal plane.

2. It is not planned in GDAAS to create new objects or create new parameters of an object from this analysis of ASM. But it is possible that new objects (component of a multiple system for example) can be detected in ASM. If simulations show that objects can be found only in ASM, the possibility of creating them will have to be examined.

2.1.7 Provider

Nol Robichon

2.1.8 Schedule

Expected algorithm delivery: July 1st 2003 for a (very) prelimary version. September 2003 for a more sophisticated one.
2.2 Fundamental algorithms (1-03-C)

2.2.1 Function

To compute the proper direction to an arbitrary source (quasar, star or solar-system object) at an arbitrary time.

2.2.2 Method

All calculations are made in a general-relativistic framework of the appropriate accuracy. Effects to be taken into account include gravitational deflection by the Sun, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune and aberration to third order in $1/c$ (that is, neglected terms are $O(c^{-4})$). Deflection caused by the moons and quadrupole gravitational fields are neglected, although these effects in exceptional circumstances may exceed $1 \mu$as. This is done by a series of transformations going from the astrometric parameters (of a star or quasar) or a barycentric ephemeris (for a solar-system source) to the proper direction (prediction mode).

The relevant inverse transformations (going from the observed proper direction to satellitocentric or barycentric coordinate direction: correction mode) are also part of the package.

The following sections describe Version 2 of the Fundamental algorithms (FA02, 2 July 2003). This is a self-contained package of 11 Fortran90 modules (see Remark 6) including ephemerides and a main program that validates the implementation of the major program elements (see Remark 7).

Although the package contains subroutines both for the prediction and correction mode, only the prediction mode is described in detail here. There are two basic subroutines in prediction mode:

- $\text{grTranA2S}$ is used for (single) stars and quasars and takes the standard six astrometric parameters as input;

- $\text{grTranE2S}$ is used for solar system sources and assumes that a barycentric ephemeris exists for any such object, accessed via a source identification number (see Remark 3).

Most of the input/output arguments are the same for $\text{grTranA2S}$ and $\text{grTranA2S}$, and they are therefore described jointly in the following sections. The names of Fortran variables are written in typewriter type.

2.2.3 Input data

- $\text{flnder}$ is a logical flag indicating whether partial derivatives are to be computed.

- $\text{accuracy}$ is a real number ($\epsilon$) indicating the desired level of accuracy in radians. Terms that are estimated to influence the proper direction by an angle $< \epsilon$ are neglected in the computations in order to gain speed. Putting $\epsilon \leq 0$ causes all
computations to be made at the highest accuracy permitted by the algorithm and/or numerical precision. Normally this should give errors < 0.1 µas. See Remark 5.

• $g$ is a real array of global parameters. Only the first element $g(1)$, containing the PPN parameter $\gamma$, is used.

• $ep0$ ($grTranA2S$ only) is the epoch $t_0$ to which the astrometric parameters refer. It is expressed in days of TCB (TBC) from the arbitrary origin JDORIGIN defined as a Julian Date.\footnote{The module GR_Constants.f90 defines JDORIGIN as JD2451545.0 (J2000). Note however that all ephemerides are defined on an absolute time scale (Julian Date) and therefore independent of JDORIGIN.}

• $a$ ($grTranA2S$ only) is a real array containing the six astrometric parameters at epoch $t_0$. These are:

$$
\begin{align*}
    a_1 &= \alpha_0 \quad \text{(right ascension at } t_0, \text{ in [rad])} \\
    a_2 &= \delta_0 \quad \text{(declination at } t_0, \text{ in [rad])} \\
    a_3 &= \pi_0 \quad \text{(parallax at } t_0, \text{ in [rad])} \\
    a_4 &= \mu_{\alpha} \ast 0 \quad \text{(proper motion in } \alpha \text{ at } t_0, \text{ times } \cos \delta_0, \text{ in [rad day}^{-1}]) \\
    a_5 &= \mu_{\delta} 0 \quad \text{(proper motion in } \delta \text{ at } t_0, \text{ in [rad day}^{-1}]) \\
    a_6 &= \mu_r 0 \quad \text{(the radial velocity parameter at } t_0, \text{ in [rad day}^{-1}])
\end{align*}
$$

(1)

where the radial-velocity parameter is defined as $\mu_r = v_r \pi/A$, $A$ being the astronomical unit. All six parameters are ‘apparent’ quantities in the sense discussed by Klioner (2003).

• $sourceID$ ($grTranE2S$ only) is an integer identifier for the source (solar-system object), see Remark 3.

• time of observation ($t_{obs}$) expressed in days of TCB (TBC) from JDORIGIN.

\subsection{2.2.4 Output data}

• $s(1:3)$ is a real array containing the CoMRS coordinates of the the proper direction to the source ($s$), i.e. a unit vector pointing towards its observable position

• $dsdg(1:3)$ is a real array containing the partial derivatives $\partial s/\partial \gamma$ of $s$ w.r.t. the global parameter $\gamma = g(1)$

• $dsda(1:3,1:6)$ ($grTranA2S$ only) is a real array containing the partial derivatives of $s$ w.r.t. the source parameters $a_1 \ldots a_6$

• $dsdx(1:3,1:3)$ ($grTranE2S$ only) is a real array containing the partial derivatives of $s$ w.r.t. the barycentric source coordinates $x_e$ (see Remark 3).
2.2.5 Data access requirements

Sources:
- READ single / ID
- WRITE none

Global parameters:
- READ all
- WRITE none

Ephemerides:
- READ single / time
- WRITE none

Minor planets orbital elements:
- READ none or single (depending on ID) / time
- WRITE none

2.2.6 Remarks

1. The algorithm uses a model of positional observations in a general-relativistic framework, based on the Parameterized Post-Newtonian (PPN) formalism, as described in Klioner (2003). Details of the implementation will be given elsewhere.

2. Barycentric ephemerides of the observer (Gaia) and of the Sun, Venus, ..., Neptune are provided through a set of self-contained subroutines in modules Ephem.*.f90.

3. For a source in the solar system, it is assumed that the barycentric coordinates \( x_e \) at the arbitrary time of light emission \( t_e \) (TCB) are provided through a call to the subroutine sourceEphem:

\[
\text{CALL sourceEphem (sourceID, datjd, xe, ierr)}
\]

Here, the input arguments sourceID (integer) and datjd (real) are the source identifier and the Julian Date of \( t_e \), while the output consists of the real array \( xe(1:3) \) with the barycentric coordinates in [m] and an error flag (ierr = 0 on normal return). In FA02 this subroutine only accepts 9 different source identifiers, viz. 0, 2, ..., 8 for the Sun, Venus, ..., Neptune, and 100 for a hypothetical source moving with constant velocity through the solar system (roughly at the distance of Jupiter). E.g. putting sourceID = 6 allows to calculate the proper direction to Saturn, taking into account, as required, the gravitational deflections from the other solar-system bodies.

4. In order to use these transformations in an iterative adjustment procedure (such as GIS), partial derivatives of the proper direction with respect to all (adjustable) model parameters are required and provided as described in the output data. However, in certain

\[\text{[References]}\]


\[5\] The analytical calculation of the partial derivatives sometimes uses a number of approximations, but the result (checked against numerical differentiation) should always be correct to at least the first few
other circumstances the derivatives are not needed and would only result in unnecessary computations. For this reason a logical flag (flder) is provided as an input argument to grTranA2S and grTranE2S. Putting flder = .FALSE. causes the computation of derivatives to be skipped and the corresponding output arrays are set to zero.

The formal accuracy aimed at in the algorithm is about 0.1 \( \mu \)as. However, for some purposes (e.g. source matching) the full accuracy may not be needed and computing time could be saved by skipping some finer details of the algorithm. For this reason, the input argument accuracy (\( \epsilon \)) is provided, the value of which determines whether any simplifying approximations will be used. In general such decisions are made conservatively, meaning that the neglected terms are usually much smaller than \( \epsilon \). For full accuracy, set accuracy \( \leq 0 \). No saving is gained by making \( \epsilon > 10^{-7} \) (20 mas).

The package FA02 contains the following Fortran90 elements:

PROGRAM GR_Test (validation program)
MODULE GR_Params (define constants)
MODULE GR_Constants (define constants)
MODULE GR_Error (simple error handling)
MODULE GR_Vectors (vector operations etc.)
MODULE GR_Transformations (the basic GR transformations)
MODULE Ephem_Gaia (ephemeris of Gaia)
MODULE Ephem_Planets (ephemerides of Sun and major planets)
MODULE Ephem_Source (ephemeris of an arbitrary source in the solar-system)
PROGRAM Ephem_Test (not used – validates Gaia and planets ephemerides separately)
MODULE Ephem_Utils (utilities for the ephemerides)

The main program (GR_test) provided with the package performs a self-validation of the implementation, including the ephemerides (thus there is no need to run Ephem_Test separately). If successful, the following should be printed to standard output:

self-test of module Ephem_Planets OK
self-test of module Ephem_Gaia OK
self-test of procedure grTranA2S OK
self-test of procedure grTranE2S OK

2.2.7 Provider

Lindegren

2.2.8 Schedule

Algorithm delivered: 2 July 2003 (FA02)

significant figures. This is sufficient in an iterative adjustment procedure, but the user should be aware of the fact that approximations are used.
2.3 Improved attitude modelling (1-04-C)

2.3.1 Function

To model the attitude as function of time in such a way that sudden rate changes (as caused e.g. by micrometeoroids) can be accommodated.

2.3.2 Method

The attitude quaternions, as function of time, are described by spline functions fitted to observed angles and rates. The required flexibility is obtained by means of adaptive knot placement, based on an analysis of rate residuals.

The spline fitting is made in one or more passes, depending on the degree of complications encountered. A first pass uses an equidistant knot sequence as in GDAAS-1, from which residuals are computed. The fit is accepted if the residuals are below a pre-defined level. Otherwise, a revised knot sequence is constructed based on an analysis of the residuals. The process is repeated until the residuals are acceptable, or the maximum number of iterations has been reached.

A preliminary version of the algorithm simply allows the knot sequence to be changed without analysis of residuals. The purpose of this version is to test the possible use of relationships between spline intervals and (e.g.) observations.

2.3.3 Input data

Input data are the same as in GDAAS-1, except that the improved instrument model has more parameters. Additional control parameters are the maximum number of iterations and thresholds for acceptable residual statistics.

2.3.4 Output data

Output data are the same as in GDAAS-1, plus diagnostic output such as: statistics of positional residuals for original (equidistant) and modified (if any) knot sequences, number of additional knots, number of multiple knots, estimated rate discontinuities, flag for OK end results.

2.3.5 Data access requirements

Elementary data:

READ batch / time
WRITE none (but update links to attitude?)
2.3 Improved attitude modelling (1-04-C)

Sources:
- READ batch / link from observation / well-behaved stars
- WRITE none

Attitude:
- READ batch / time
- WRITE batch / those read / update

Calibration:
- READ batch / time / ASM and AF
- WRITE none

Global parameters:
- READ all
- WRITE none

Ephemerides:
- READ batch / time
- WRITE none

2.3.6 Remarks

1. Much of the diagnostics are well suited for graphical display (plots of residuals versus time, etc). This is also a case where visual inspection and the possibility of human interaction may be both feasible and desirable.

2.3.7 Provider

Preliminary version: Lindegren
Update: van Leeuwen

2.3.8 Schedule

Expected algorithm delivery:
July 2003 (preliminary)
April 2004 (update)
2.4 Improved instrument modelling (1-05-C)

2.4.1 Function

To calibrate the line spread function (LSF) in the AF, BBP and MBP, for one particular calibration unit (one CCD in a given time interval).

2.4.2 Method

Robust weighted least-squares fitting of suitable functions to samples. Normal equations are accumulated by several instances of the accumulation module, and later combined and solved in the solution module.

Only well-behaved sources are used, i.e., they must show no sign of duplicity or variability. For these sources only well-behaved observations are used, i.e., they must not be close to saturating the CCD and must not have been disturbed by cosmics or by objects from the other field of view. The selection of well-behaved observations is not part of the present algorithm.

Each instance of the accumulation module handles only one particular observation window configuration (number of samples, width of each sample) and only one particular set of photometric small scale calibration coefficients. Observations made in other window sizes or in other photometric small scale calibration units will be accumulated in other instances, and everything is eventually combined before the next iteration of the solution.

A certain fraction (\(\sim 0.1–0.5\)) of the signal, to be chosen at run time, is modelled by a fixed Cauchy distribution, and the remaining signal by ‘bi-quartic B-splines’ with a knot separation of 0.5 pixels, as described in GAIA–LL–046.

The LSF is decomposed into a number (one or more) of quasi-monochromatic response functions representing different photometric bands. The photometric bands should probably be observed in the same instrument as the one being calibrated (BBP bands for Astro CCDs; and MBP bands for Spectro CCDs). For the AF CCDs, the LSF is then decomposed into 5 quasi-monochromatic response functions, weighted according to the normalised BBP standard fluxes of each source. For the BBP and MBP CCDs, a single quasi-monochromatic response function may be sufficient. The actual choice of bands is a decision for the implementation.

The observation equations are given by (15) in the Appendix.

2.4.3 Input data

Input data for the accumulation module are:

- instrument (Astro or Spectro) where the CCD is located
- number of quasi-monochromatic response functions, \(n_{\text{band}}\), used to model the LSF
2.4 Improved instrument modelling (1-05-C)

- wavelengths, $\lambda_i$, $i = 1 \ldots n_{\text{band}}$, for each band
- fraction of the signal modelled as a Cauchy distribution
- number of knots, $n_{\text{knot}}$, for each quasi-monochromatic response function
- current LSF for this calibration unit
- configuration of the observation window (width of each sample)
- photometric calibration parameters
- chromaticity parameters (only relevant for AF)
- data for each of the selected observations
  - individual samples
  - background estimate for a unit width sample
  - estimated centroid position [pixels]
  - photometric large and small scale calibration unit
  - standard fluxes needed for modelling the observed flux

Input data for the solution module are:

- fraction of the signal modelled as a Cauchy distribution
- number of quasi-monochromatic response functions, $n_{\text{band}}$, used to model the LSF
- number of knots, $n_{\text{knot}}$, for each quasi-monochromatic response function
- the accumulated equations, $n_{\text{band}}n_{\text{knot}}(n_{\text{band}}n_{\text{knot}} + 3)/2$ numbers
- current LSF for this calibration unit

2.4.4 Output data

Output data from the accumulation module are:

- the accumulated equations, $n_{\text{band}}n_{\text{knot}}(n_{\text{band}}n_{\text{knot}} + 3)/2$ numbers

Output data from the solution module are:

- updated LSF parameters, including uncertainties
- coordinate origins, $u_0$, for each quasi-monochromatic response function
2.4.5 Data access requirements

Observations:
  READ batch / time and CCD
  WRITE none

Sources:
  READ batch / link from observation / well-behaved stars
  WRITE none

Calibration:
  READ batch / time and CCD / LSF and photometric calibration
  WRITE batch / those read / updated LSF

2.4.6 Remarks

1. See Appendix A and GAIA–LL–046 for the calibration model.

2. For GDAAS-2 only the (along-scan) LSF is updated. The point spread function, PSF, is modelled from the LSF and an assumed across-scan LSF. The LSF is determined relative to the centroid of the image; thus essentially a local process. The calibration of the PSF would be a significantly different (and more complex) algorithm, using also the astrometric source data, attitude and geometric calibration to centre the PSF.

3. A better estimation of the LSF (and PSF) would use the predicted transit time instead of the observed centroid. This will be possible in the more advanced algorithm mentioned in Remark 2.

4. Stars of intermediate brightness ($G \simeq 12–16$) are probably best suited for this calibration, since wider windows are transmitted for these; but fainter stars are also accepted.

5. The information as to which stars are ‘well-behaved’ could come from the discrete source classifier. In any case, the present algorithm assumes that all observations presented to it are well-behaved.

6. The gain for the CCD amplifier is assumed to be $1 \, e^-/\text{adu}$. If needed, the gain could be an input parameter for the accumulation module.

2.4.7 Test data

A program Tdat (tdat105.f) generates sets of data for testing algorithm 1-05-C. The data sets are not meant to be realistic, but merely allowing a test of the validity of the code. There are test files for a model using one photometric band (for BBP and MBP) and for a model using five bands (for AF). There are files with noise-free observations, expected to reproduce the input LSF very closely, and files with some light, primitive noise added. In all cases, each observation contains six samples, and the core of the LSF is defined using nine knots. Files are generated with 10, 25 and 999 observations.
Another program, Test (tes105.f), accepts one of the output files from Tdat as input, finds an improved LSF and compares it with the previous LSF.

2.4.8 Provider

Fabricius

2.4.9 Schedule

Expected algorithm delivery: July 2003
2.5 Detailed geometrical calibration (1-06-C)

2.5.1 Function

To calibrate the geometric characteristics of the Astro instruments (ASM, AF and BBP) on all spatial and temporal scales.

2.5.2 Method

Robust weighted least-squares fitting (reducing to robust averaging for a single parameter). Normal equations are separately accumulated in different bins according to CCD number and source colour, and solved after the required time interval.

2.5.3 Input data

- time interval and range of CCDs
- centroid positions (in pixel coordinates) for well-behaved sources observed in the interval
- geometrical calibrations relevant to the time interval and CCDs
- attitude parameters for the time interval
- source parameters
- global parameters

2.5.4 Output data

- improved geometric calibration parameters, including uncertainties

2.5.5 Data access requirements

Elementary data:
```
READ batch / time and CCD
WRITE none
```

Sources:
```
READ batch / link from observation / well-behaved stars
WRITE none
```

Attitude:
```
READ batch / time
WRITE none
```
2.5 Detailed geometrical calibration (I-06-C)

Calibration:
READ batch / time and CCD
WRITE batch / those read / update

Global parameters:
READ all
WRITE none

Ephemerides:
READ batch / time
WRITE none

2.5.6 Remarks
1. See Appendix A for the calibration model.

2.5.7 Provider
Bastian

2.5.8 Schedule
Expected algorithm delivery: April 2004
2.6 Chromaticity calibration (1-07-C)

2.6.1 Function

To calibrate chromaticity in AF on all spatial and temporal scales.

2.6.2 Method

Robust weighted least-squares fitting (reducing to robust averaging for a single parameter). Normal equations are separately accumulated in different bins according to CCD number and source colour, and solved after the required time interval.

2.6.3 Input data

- time interval and range of CCDs
- centroid positions (in pixel coordinates) for well-behaved sources observed in the interval
- geometrical calibrations relevant to the time interval and CCDs
- attitude parameters for the time interval
- source parameters
- global parameters

2.6.4 Output data

- improved chromaticity parameters, including uncertainties

2.6.5 Data access requirements

Elementary data:
READ batch / time and CCD
WRITE none

Sources:
READ batch / link from observation / well-behaved stars
WRITE none

Attitude:
READ batch / time
WRITE none
2.6 Chromaticity calibration (1-07-C)

Calibration:
READ batch / time
WRITE batch / those read / update

Global parameters:
READ all
WRITE none

Ephemerides:
READ batch / time
WRITE none

2.6.6 Remarks

1. See Appendix A for the calibration model.

2.6.7 Provider

Lindegren

2.6.8 Schedule

Expected algorithm delivery:
July 2003 (preliminary)
April 2004 (update)
2.7 Improved CCD calibration (1-08-C)

2.7.1 Function

To calibrate the non-linear part of the CCD behaviour (due to CTI effects, near saturation, etc.), as function of time, source brightness, background, and other relevant factors (TBD).

2.7.2 Method

TBD

2.7.3 Input data

TBD

2.7.4 Output data

TBD

2.7.5 Data access requirements

TBD

2.7.6 Remarks

2.7.7 Provider

Holland

2.7.8 Schedule

Expected algorithm delivery:
April 2004 (preliminary)
September 2004 (update)
2.8 Additional global parameters (1-09-S)

2.8.1 Function

To determine arbitrary global parameters, as specified by the user. (A specific choice for testing is proposed in Remark 2 below.)

2.8.2 Method

Robust weighted least-squares analysis of positional (along-scan) residuals for well-behaved sources.

2.8.3 Input data

- specification of global parameters
- time interval and range of CCDs considered
- type of sources used

2.8.4 Output data

- estimated global parameters, with uncertainties

2.8.5 Data access requirements

Elementary data:
- READ batch / time and for the selected sources
- WRITE none

Sources:
- READ batch / source classification / well-behaved, unambiguous quasars
- WRITE none

Attitude:
- READ batch / time
- WRITE none

Calibration:
- READ batch / time
- WRITE none

Global parameters:
- READ all
- WRITE all / update
Ephemerides:
  READ batch / time
  WRITE none

2.8.6 Remarks

1. In general, this task should allow simple investigation of questions like: ‘Is there a
global 4th harmonic variation of the basic angle with a fixed phase relation to the solar
direction?’

2. However, for the purpose of GDAAS-2 it is proposed to simply implement a few specific
additional global parameters, viz. the inertial rotation $\omega$ and the galactic acceleration $a$
discussed in the GAIA Study Report, Sect. 1.8.11. The determination of these six
additional aglobal parameters requires the selection of a special subset of the sources,
‘well-behaved and unambiguous quasars’ (i.e. the contamination by stars must be small).

3. It is assumed that the processing for the selected sample of sources is made in batches
corresponding to contiguous time intervals (although other schemes could also be used,
which would modify the data access requirements).

2.8.7 Provider

Mignard

2.8.8 Schedule

Expected algorithm delivery: April 2004
2.9  Visual double star analysis (1-10-S)

2.9.1  Function

To determine the astrometric, (BBP) photometric and relative astrometric parameters for any source classified as a resolved double or multiple star.

2.9.2  Method

Robust weighted least-squares fitting of astrometric, photometric and system parameters to all available sample data (ASM, AF and BBP) of the source.

2.9.3  Input data

- a reference point \((\alpha, \delta)\) and angular radius \((\rho)\), defining a small area on the sky to be analysed
- number of components \((n)\) to be fitted and the type of source model (e.g. fixed, optical, linear, ...)

2.9.4  Output data

- quality of the fit
- for each of the \(n\) components, a set of astrometric and photometric parameters with covariances

2.9.5  Data access requirements

Observations:
  READ batch / link from sources / ASM, AF and BBP data
  WRITE none

Sources:
  READ single or few / position and radius
  WRITE single or few / those read / update + insert new components

Attitude:
  READ batch / link from observations
  WRITE none

Calibration:
  READ batch / link from observations
  WRITE none
Global parameters:
- READ all
- WRITE none

Ephemerides:
- READ all
- WRITE none

2.9.6 Remarks

1. The sampling strategy in Astro for resolved and partially resolved objects has not been fixed. Generally speaking, if the components of a pair are sufficiently separated they are observed as two independent sources, each getting a centred window of $6[1] \times 1[10]$ samples (TBC). If they are closer, a single big window ($16[1] \times 1[14]$ samples, TBC) is instead placed at the geometrical centre of the pair. The algorithm should be flexible enough to cope with (almost) arbitrary window/sample sizes.

2. If a pair is barely resolved, both components are always observed jointly and the pair will have a single ID in the data base. (Roughly, this happens for separations below 0.3 arcsec).

3. If the pair is consistently observed as two sources, they will have separate IDs and the algorithm needs to combine all the Astro data for both sources. (This happens for separations above 1.3 arcsec.)

4. At intermediate separations the pair will sometimes be observed as two sources (with separate IDs), sometimes as one source (possibly using an extended window), which may be matched to either of the two IDs or to a third. A faint companion can sometimes go undetected in some scans.

5. Because of these various complications, the task must begin by identifying all relevant observations.

2.9.7 Provider

Söderhjelm

2.9.8 Schedule

Expected algorithm delivery:
September 2003 (preliminary)
April 2004 (update)
2.10 Astrometric Binary star analysis (1-11-S)

2.10.1 Function

For any source classified as unresolved double or multiple, this task performs a periodogram analysis of positional residuals and attempts to describe the photocentric motion by a Keplerian orbit.

2.10.2 Method

It is assumed that the single star fit has been performed. The best single star solution is therefore available together with the (abscissa) residuals as well as the partial derivatives of the abscissa with respect to that single star solution (six parameters).

Owing to the continuous change of the scanning direction, the abscissa residuals cannot be directly used to perform a period analysis. However, for any triplet \((e, P, T)\), deriving the best fitting orbital solution (six astrometric parameters + the four Thiele-Innes constants) is straightforward as the solution of an over-determined system of linear equations. One can even assume that the best fitting period \((P)\) is not tied too strongly to the eccentricity \((e)\) and thus limit the investigation to circular orbits. If so, the periastron time \((T)\) is set to any time. One therefore performs a period analysis of the chi square where \(P\) is the only parameter free, the other ones are solutions of the chi-square minimization using a Singular Value Decomposition algorithm. In order to avoid problems of aliasing caused by the scanning law, version 1.0 of this module does not seek periods shorter than ten days.

Once the best period and the subsequent parameters are guessed, they are all adopted as the starting point of a local minimization where the thirteen parameters are fitted together. The Levenberg-Marquardt method was selected for this section where one no longer takes advantage of the linearity of the model over the first ten parameters.

2.10.3 Input data

The single star solution is required and is passed to the constructor of the AstromBinary class:

- singlesol: the six parameters of the single star solution;
- \(n\): the number of observations;
- obstime: the time of the \(n\) observations;
- parder: the \(n \times 6\) matrix of the partial derivatives of the abscissa with respect to the six parameters of the single star solution;
- res: the \(n\) abscissa residuals;
- weight: the weight matrix. For the time being (Vers. 1.0), only the diagonal elements are used, i.e. a 0 correlation between the observations is assumed.
2.10.4 Output data

The method FitOrbit returns the 13 parameters together with their covariance matrix, the chi square and the goodness of fit:

- FullParam: the thirteen parameters of the solution;
- CovMat: the covariance matrix of the parameters;
- QualityInd: a vector of two components with the final chi square and the goodness of fit.

2.10.5 Data access requirements

Besides the data required by the constructor and returned by FitOrbit, this module does not need any access to the database.

2.10.6 Remark

All angles are reckoned in radians. The orbital period and the periastron time are in the same unit as other time and duration (observation time). $m_{\text{obsTime}}$ is assumed to be absolute. If the time interval covered by the observations is larger than 10, the code figures out that the times are reckoned in days (in years otherwise).

2.10.7 Test data

2.10.8 Validation data

A file with 50 test cases is supplied. It includes the 6 initial parameters, the observation times, partial derivatives, residuals and observation error. It also includes the 13 parameters of the best astrometric orbital solution.

2.10.9 Provider

D. Pourbaix.

2.10.10 Schedule

Expected algorithm delivery: July 1st 2003
2.11 Astrometric planet detection (1-12-S)

2.11.1 Function

To detect planetary companions astrometrically and to determine the Keplerian elements of the star due to each detected companion.

2.11.2 Method

Periodogram analysis of positional residuals; identification of significant periods; robust weighted least-squares fitting of orbital parameters. Alternatively, multi-dimensional global minimisation by non-classical methods, or hybrid techniques.

2.11.3 Input data

- source ID

2.11.4 Output data

- number of planetary companions found $n_P$
- astrometric parameters for the mass centre of the binary (the usual six parameters)
- for each planet $(j = 1 \ldots n_P)$, seven orbital parameters, e.g. $a_j$, $P_j$, $\tau_j$, $\omega_j$, $\Omega_j$, $e_j$, $i_j$
- error estimates of the above parameters
- value of the objective function (e.g. $\chi^2$) and its significance

2.11.5 Data access requirements

Elementary data:
READ batch / link from source / AF and BBP data
WRITE none

Sources:
READ single / ID
WRITE single / the one read / update

Attitude:
READ batch / link from observations
WRITE none

Calibration:
READ batch / link from observations
WRITE none
2.11.6 Remark

1. The method, input/output and data access requirements could essentially be the same as for *Astrometric binary star analysis* (1-11-S), except that the planet detection in principle allows more than one companion per star.

2.11.7 Provider

Lattanzi

2.11.8 Schedule

Expected algorithm delivery: July 2003
2.12 Astrometry quick-look analysis (1-13-Q)

2.12.1 Function

To perform a first-look assessment of the geometric stability of the instrument, smoothness of the attitude and proper astrometric working of the spacecraft, by means of a Hipparcos-type great-circle reduction.

2.12.2 Method

The basic method will be quite similar to the great-circle reduction of Hipparcos. Re-use of software parts from the Hipparcos reductions should be possible.

2.12.3 Input data

- definition of the great-circle interval to be treated (a largely contiguous period of about 12–24 hours, without significant onboard events in between)

- elementary observations (with uncertainties), for well-behaved sources, for the great-circle interval

- current geometric calibration parameters for the great-circle interval

- current attitude parameters for the same period of time, optimally from the mainstream attitude processing (if this can be performed in close pursuit of the telemetry arrival)

- current source parameters

- current global parameters

2.12.4 Output data

- improved large-scale along-scan geometric parameters (and perhaps across-scan geometric parameters for the ASMs)

- improved attitude parameters along scan

- improved source positions along scan

- diagnostics on the geometric stability of the instrument, smoothness of the attitude and proper astrometric working of the spacecraft
2.12.5 Data access requirements

Elementary data:
- READ batch / time and CCD
- WRITE none

Sources:
- READ batch / link from observation / well-behaved stars
- WRITE none

Attitude:
- READ batch / time
- WRITE none

Calibration:
- READ batch / time
- WRITE none

Global parameters:
- READ all
- WRITE none

Ephemerides:
- READ batch / time
- WRITE none

2.12.6 Remarks

1. The results of this reduction are used to derive diagnostics on the geometric stability of the instrument, smoothness of the attitude and proper astrometric working of the spacecraft. The main criterion will be the residues from the great-circle adjustment. Another important one is the comparison of the newly derived calibration with the ones from preceding great-circle intervals.

2. In principle the results include updated estimates of the attitude and some calibration parameters, which could be written back to the database. However, it is not clear whether this is advisable (especially since the update can be only partial, e.g. no transverse attitude). For the time being it is assumed that the results are not used to update the database.

Rationale for the astrometry quick-look analysis (by U. Bastian):

On one hand, the global iterative solution (GIS) needs of the order of 0.5–1 year of mission data to get the first results with a precision approaching the mission goals (i.e. tens of $\mu$as). On the other hand, any quick-look inspection of the mission data can go only to the precision level of the source parameters and calibration parameters available. Without any further measures, this would be of the order of tens of mas the first few months of the mission. Thus, any disturbance of the measurements even at the level of hundreds of $\mu$as would remain undetected.

This is unsatisfactory, since technically possible countermeasures could not be taken, and a poten-
tially significant portion of the mission data could be lost.

Example causes for disturbances at the 100 $\mu$as level are instabilities of some FEEP, thermal clicks in the spacecraft structure, erratic heat sources deforming the payload on short time scales, erratic sources of magnetic moment interacting with the interplanetary magnetic field, instability of the instrument clock. Although GAIA is being specifically designed to avoid all this, it would be good just to know. Switching to a redundant FEEP or clock, identification and avoidance of varying heat flows by operational modifications etc. could then be undertaken.

A Hipparcos-type great-circle reduction is the only known way to see the stability of the payload geometry and the smoothness of the attitude quasi-instantaneously, down to the intrinsic precision of the measurements, and without the need of any external reference of similar precision.

These are the reasons why such a reduction should be foreseen at least during the first 6–12 months of the scientific mission. Later on, when a first global solution has been achieved, the necessary first-look diagnostics on the current astrometric mission performance can in principle be derived from the residues of attitude processing — provided that attitude processing is constantly achieved in close pursuit of the telemetry arrival. Alternatively, the great-circle reduction can be continued throughout the entire mission, if this is operationally simpler.

In addition to its role in the scientific data reduction proper, the great-circle reduction will be an important tool for the commissioning phase of the mission.

During the main mission, the great-circle reduction and first-look diagnostics will be run as soon as a sufficiently large and sufficiently contiguous stretch of new telemetry has become available. The great-circle intervals will be defined on the basis of satellite and payload status logs, either manually or by some automatic procedure. Depending on the outcome they may need to be modified, the great-circle reduction and first-look diagnostics to be repeated on the revised intervals.

2.12.7 Provider

Bastian

2.12.8 Schedule

Expected algorithm delivery: September 2004
2.13 Photometric calibration (2-01-C)

2.13.1 Function

(1) To determine the standard fluxes $g_{im}$ for a subset of well-behaved stars;

(2) To calibrate the photometric characteristics of each CCD on all spatial and temporal scales, using the subset of well-behaved stars. This means to calibrate the differences in response between all the CCDs, and their individual columns, covering a specific photometric band. The algorithm should have sufficient flexibility to allow calibration of the ASM, AF, BBP, MBSM, MBP and the RVSM.

2.13.2 Method

The determination of standard fluxes is either carried out as a least squares estimation for several bands at a time, or as a robust mean estimation for an individual band if any auxiliary fluxes can be assumed known. It is a decision for the implementation to choose which method to use in each case. A suggestion could be:

(1) Robust weighted least-squares estimation for one source at a time of $g_{im}$ from (22), either solving for all the Astro bands or for all the MBP bands;

(2) Robust weighted mean estimation for one source at a time of $g_{im}$ from (22), for the RVSM bands, where MBP fluxes can be assumed known;

(3) Robust weighted least-squares estimation of the parameters $A_{km}$ and $a_{km}$ from (23) and (24), for one (large scale or small scale) calibration unit at a time.

Only well-behaved sources are used, i.e., they must show no sign of duplicity or variability. Among these sources, only well-behaved observations are used, i.e., they must show no saturation and must not have been disturbed by cosmosics or by sources in the other field of view. Selecting the well-behaved sources and observations is not part of the present algorithm.

The estimation of calibration parameters involves a large data set and is therefore divided into two tasks. (A) An accumulation module will accumulate normal equations for a subset of the data; (B) a solution module will solve the equations once all data sets have been processed. Steps (A) and (B) will need a few iterations.

2.13.3 Input data

Input data for least squares estimation of $g_{im}$ are:

- photometric calibration parameters
- current standard fluxes for the source
- estimated (observed) fluxes, $f_l$, including uncertainties
Input data for the estimation of calibration parameters are:

- number of bands used for modelling the observed flux
- standard fluxes in these bands, including uncertainties, for well-behaved stars observed in the calibration unit
- estimated (observed) fluxes, $f_i$, including uncertainties, for well-behaved observations of these sources
- current photometric calibration parameters, relevant for these observations

### 2.13.4 Output data

- estimated standard fluxes $g_{im}$, including uncertainties
- estimated large-scale photometric parameters $A_{km}$, including uncertainties
- estimated small-scale photometric parameters $a_{km}$, including uncertainties

### 2.13.5 Data access requirements

**Elementary data:**
- READ batch / selected sources / well-behaved observations
- WRITE none

**Sources:**
- READ batch / position (HTM node) / well-behaved stars
- WRITE batch / same as read / update

**Calibration:**
- READ batch / CCD
- WRITE batch / same as read / update

### 2.13.6 Remarks

1. See Appendix A for the calibration model.
2. The photometric raw data treatment is part of the algorithm for estimating the centroid position, flux and background (see UB-GDAAS-TN-008).
3. The standard magnitudes are determined for one star at a time.
4. The calibration coefficients are determined for one calibration unit at a time. Normalisation of calibration coefficients must be carried out every time the calibration coefficients in the data base have been updated.
5. The equations for the calibration coefficients may be accumulated concurrently on several nodes, and later combined for solving and normalisation. Data access could be according to CCD and time.

6. The information as to which stars are ‘well-behaved’ could come from the discrete source classifier. In any case, that is not part of the present algorithm.

7. Care should be taken by the calling program, that the ‘well-behaved’ flag is not updated while the photometric calibration is in progress.

2.13.7 Test data

A set of Fortran 77 programmes was written that call the subroutines in the Fortran library for algorithm 2-01-C. These programmes only carry out tests of the individual subroutines, i.e., no attempt was made at trying to run the algorithms in an integrated way by iterating between them.

The tests require simulated data which consist of:

1. representative stellar fluxes
2. a model of the calibration parameter values

The stellar fluxes were taken from the simulated photometry provided by the Barcelona group (see document UB-PWG-013). The simulated fluxes are provided for a wide range of stellar types for the $G$-band, the RVSM band, the BBP bands in the 2B and 3B systems and the MBP bands in the 1X and 2F systems. The observed fluxes $f_l$ for the test sources are then modelled according to (18). For the ASM and AF CCDs $m = 0 \ldots 5$, where $m = 0$ stands for the $G$-band and $m = 1 \ldots 5$ for the BBP bands. For the BBP CCDs only the five BBP bands ($m = 1 \ldots 5$) are used to model the flux. In the case of the Spectro instrument $m = 0 \ldots 5$, where $m = 0$ stands for the $G_m$ band (MBSM) or one of the MBP bands (note that the simulated RVSM flux was used to generated the $G_m$ fluxes). Again, $m = 1 \ldots 5$ stands for the BBP bands.

A special purpose Fortran library was written that generates the values of the calibration parameters $A_{km}$ and $a_{km}$. The values are based on very simple functional forms that describe the variations of $A_{km}$ as a function of CCD and time, and of $a_{km}$ as a function of pixel column and time. The time scale over which the $A_{km}$ vary is 1 day and for the $a_{km}$ this is 90 days (3 months).

For the large scale parameters the calibration unit numbers $k$ are generated as follows. The CCDs in the focal plane are labelled according to their across-scan coordinate (or index) $i_{ac}$ and their along-scan coordinate $j_{al}$. The time $t_{lsc}$ is tagged in units of 1 day, where ‘lsc’ stands for ‘large scale’. The large scale calibration unit number $k$ is then given by:

$$k = N_{al}(i_{ac} - 1) + j_{al} + (t_{lsc} - 1)N_{al}N_{ac},$$

where $N_{al}$ and $N_{ac}$ stand for the number of CCDs in the along-scan and across-scan
For the small-scale calibration unit numbers $\kappa$ a similar equation holds:

$$\kappa = 1 + (N_{al}(i_{ac} - 1) + j_{al} - 1)n_c + (t_{ssc} - 1)N_{al}N_{ac}n_c,$$

where $n_c$ stands for the number of pixel columns in the across-scan direction on each CCD, and $t_{ssc}$ is now the time in units of 90 days. In practice one has to deal with the fact that, for example, in Astro there are two telescopes that see different CCD rows. This is all taken care of in the library subroutines.

For the large scale parameters a product of two polynomials in $x$ and $y$ is used to model the variation of their values over the focal plane. The $x$-coordinate coincides with $i_{ac}$ and $y$ with $j_{al}$. The value of $A_{km}$ at CCD $(i_{ac}, j_{al})$ is then $\prod_p(i_{ac} - x_p)\prod_q(j_{al} - y_q)$, where $x_p$ and $y_q$ are the zero points of the polynomials. For $a_{km}$ the value at pixel column $c$ within a CCD is given by $\prod_p(c - x_p)$, where $x$ now stands for the coordinate along the pixel column positions. The variation in time is described by a sinusoidal function that acts as a multiplication factor applied to the spatial variations described above.

For each calibration unit $k$ or $\kappa$ there are $m$ calibration parameters, where $m$ depends on the number of photometric bands that are used to model the flux for that calibration unit. The normalisation conditions (20) and (21) for $A_{km}$ and $a_{km}$ are taken into account.

The simulated stellar fluxes and calibration parameters together then provide a means of emulating the sort of data that is expected to come from the database containing the GAIA observations. This leads to the following tests.

- The least squares estimation of the $g_{im}$ is tested by simulating the observed fluxes $f_l$ for one source observed for a number of transits across the Astro and Spectro focal planes. The calibration parameters are assumed to be known.

- The accumulation of the equations for $A_{km}$ and $a_{km}$ and their subsequent solution is tested by simulating the observed fluxes $f_l$ for many different sources in the same band for the calibration units $k$ and $\kappa$.

- The normalisation of the calibration parameters is tested by simply generating a set of values for $A_{km}$ and $a_{km}$ for a given band $m$ and passing those to the relevant subroutines.

### 2.13.8 Providers

Fabricius/Brown

### 2.13.9 Schedule

Expected algorithm delivery: July 2003
2.14 Imaging analysis (2-02-S)

2.14.1 Function

To analyse the extended Astro windows (from AF11 and ASM) from the whole mission in order to map the surroundings of detected sources, in particular to find additional faint \((G > 20)\), potentially disturbing sources.

2.14.2 Method

The reconstruction of point sources could use standard deconvolution techniques: a dirty map produced by superposing all scans (accurately positioned by means of the attitude and geometric calibration) is deconvolved using the known response to a point source (from scanning pattern and photometric/LSF calibrations).

2.14.3 Input data

- source ID
- range of CCDs considered
- source parameters
- all raw observations (samples) of this source in these CCDs
- corresponding attitude data
- corresponding calibration data
- global parameters
- ephemerides

2.14.4 Output data

- map of a field around the source
- list of detected point sources in the field, with estimated positions and fluxes

2.14.5 Data access requirements

Observations:
- READ batch / source ID / selected CCDs
- WRITE none

Sources:
- READ single / ID
- WRITE single / same as read / update
2.14 Imaging analysis (2-02-S)

Attitude:
   READ batch / link from source
   WRITE none

Calibration:
   READ batch / selected CCDs
   WRITE none

Global parameters:
   READ all
   WRITE none

Ephemerides:
   READ batch / link from source
   WRITE none

2.14.6 Remarks

2.14.7 Provider

Bijaoui

2.14.8 Schedule

TBD
2.15 Variability analysis (2-03-S)

2.15.1 Function

To detect variability, and for detected variables make simple lightcurve analysis and classification.

2.15.2 Method

Analysis of calibrated photometric data, expressed in the standard (instrument) system, for a specified range of CCDs. Details TBD.

2.15.3 Input data

- source ID
- range of CCDs considered
- source parameters
- corresponding elementary data
- corresponding photometric calibration data
- satellite ephemerides (for correction to barycentre)

2.15.4 Output data

- variability class (0 = no variability detected)
- variability parameters (if class ≠ 0)
- lightcurve

2.15.5 Data access requirements

Elementary data:
  READ batch / source ID / selected CCDs
  WRITE none

Sources:
  READ single / ID
  WRITE single / same as read / update

Calibration:
  READ batch / selected CCDs / photometric
  WRITE none
2.15 Variability analysis (2-03-S)

Ephemerides:
  READ batch / link from source
  WRITE none

2.15.6 Remarks

2.15.7 Provider

Eyer

2.15.8 Schedule

Expected algorithm delivery: September 2004
2.16 Science alerts (2-04-Q)

2.16.1 Function

To raise an alert when a possible supernova or microlensing event is detected photometrically.

2.16.2 Method

Analysis of data for a newly created source in relation to observing history of that region of the sky (see Remarks).

2.16.3 Input data

2.16.4 Output data

2.16.5 Data access requirements

2.16.6 Remarks

Tentatively, a ‘sky history’ database needs to be implemented, in order to make it possible to determine whether a newly discovered source should have been detected before, had it been there. It might consist of a coarse (few arcmin) pixel map of the whole sky (using HTM?) with statistics about the number of scans or similar. Some caution is needed in interpreting the sky history since some detected sources are not observed, especially in crowded regions.

2.16.7 Provider

Wyn Evans

2.16.8 Schedule

Expected algorithm delivery: April 2004
3  Radial Velocity

3.1  Radial velocity source detection (3-01-C)

3.1.1  Function

To detect a source in the RVSM, perform the source matching and to determine sources and background contaminating the spectrum.

3.1.2  Method

(1) Source detection is made with the same algorithm as for the other sky mappers. This is not covered by the description below. (2) Source matching is made according to algorithm 1-01-C. (3) Based on the known window positioning in RVS, other sources with overlapping spectra (in a particular transit) can be identified, and links inserted in the database. (4) The background is estimated directly from the RVSM samples.

3.1.3  Input data

- search radius for source matching
- a list of detections in the RVSM
- a list of sources in the database

3.1.4  Output data

- for each detection, its matching source in the database; alternatively, data for a new source to be inserted in the database
- a list of contaminating sources
- the estimated background

3.1.5  Data access requirements

Elementary data:  
READ batch / time and CCD / RVSM only  
WRITE batch / those read / links updated  

Sources:  
READ batch / position  
WRITE batch / subset of those read + new sources / update and insert
3.1 Radial velocity source detection (3-01-C)

Attitude:
- READ batch / time
- WRITE none

Calibration:
- READ batch / time and CCD / RVSM only
- WRITE none

Global parameters:
- READ all
- WRITE none

Ephemerides:
- READ batch / time
- WRITE none

3.1.6 Remarks

3.1.7 Provider

Arenou

3.1.8 Schedule

Expected algorithm delivery: April 2004
3.2 Radial velocity wavelength calibration (3-02-C)

3.2.1 Function

To calibrate the zero point of the (barycentric) radial velocity (or wavelength) scale as function of template, source parameters, position in the field, and time. From this calibration it will be possible to compute the wavelength, in the barycentric frame, of any sample in the RVS field, assuming that the signal comes from a specific source.

3.2.2 Method

For a given calibration model (e.g., the functional representation of wavelength as function of field position), the calibration parameters should be such that (1) the resulting radial-velocity variance of bona fide constant stars is minimised; (2) that the mean radial velocity for a subset of radial velocity standards equals the mean value of their standard velocities (or, equivalently, that the mean difference between computed and standard values is zero).

3.2.3 Input data

- time interval and CCDs considered
- current wavelength calibration
- observed barycentric radial velocities (epoch values) for a number of bona fide constant stars, including some radial-velocity standards
- the times and field (or pixel) coordinates of the observations
- standard velocities for the subset of standard stars

3.2.4 Output data

- updated wavelength calibration

3.2.5 Data access requirements

Elementary data:

READ batch / link from sources / selected time and CCDs (in RVS)
WRITE none

Sources:

READ batch / ID / well-behaved + standard stars
WRITE none

Attitude:

READ batch / link from sources
WRITE none
Calibration:
   READ batch / selected CCDs (in RVS)
   WRITE batch / same as read / update

Global parameters:
   READ all
   WRITE none

Ephemerides:
   READ batch / link from sources
   WRITE none

3.2.6 Remarks

Ultimately, the zero point calibration must depend on radial-velocity standard stars. The variation across the field and any (long-term) time variation can be determined from multiple observations of the same objects, which in the mean give constant velocities. The transfer of zero point between stars could provisionally use the position obtained in the RVSM (‘raw’ radial velocities), but for maximum precision the transfer should in the end make use of the accurate astrometry and attitude available from the GIS.

3.2.7 Provider

Munari

3.2.8 Schedule

Expected algorithm delivery: April 2004
3.3 Radial velocity cross-correlation (3-03-S)

1- Introduction

This README presents briefly the function and implementation of the "Radial Velocity cross-correlation" (RVCC) algorithm (3-03-S in GAIA-LL-044). This readme will be complemented/replaced by the end of July by a more detailed ESA report. The status of the development of the RVCC algorithm is presented below in Sect. 2.4: "development status".

2- General presentation

2.1- Function

The function of the present algorithm is to derive the transit or mission radial velocity of a star (hereafter referred to as source) with respect to the solar system barycenter reference frame. The radial velocity is obtained by cross-correlating the source spectrum with a synthetic template spectrum of appropriate atmospheric parameters.

2.2- Input data

* The source "transit" or "mission" spectrum. It is assumed that:
  - the wavelength of the spectrum are expressed in the solar system barycenter reference frame
  - the source spectrum is already calibrated (instrument, wavelength) and cleaned from contaminating sources.

* The atmospheric parameters of the source: effective temperature, surface gravity and metallicity. They are used to select the appropriate template.

* The radial velocity zero point correction. This correction is assumed to take into account "all" effects (e.g. gravitational redshift), except the transformation from the satellite reference frame to the solar system barycenter reference frame which is assumed already taken into account in the wavelength calibration.
  Note: The way to handle the radial velocity zero point correction and the reference frame should be reconsidered at a later stage in the more general framework of the "spectroscopic global iterative solution" and the RV shell processing strategy.

* A library of synthetic reference templates and the list of the spectra contained in the library.
2.3- Output data

* The radial velocity of the source with respect to the solar system barycenter reference frame. The units used for the radial velocity are km/s (to be modified according to GDAAS conventions).

2.4- Development status

Simple tests have been performed on the RVCC algorithm. So far the algorithm shows the expected performances.

Two tasks remained to perform before a possible integration of the RVCC algorithm into GDAAS:

* Comments should be added in the java code.
* More elaborated tests should be performed.

An estimation of the radial velocity error should also be added.

The present version of the RVCC algorithm is posted today because it shows:

* the input/output data
* the architecture of the algorithm: class/methods (no significant change of the architecture is presently foreseen).

An updated/tested version of the the RVCC algorithm (including comments) will be proposed to GDAAS by the end of July.

In addition to the algorithm two libraries of spectra are provided:

* 49 source spectra to be used by GASS to simulate RVS observations.
* 1237 synthetic template spectra.

3- Method

The RVCC algorithm performs the following tasks:

* Select and read the synthetic template which is the most similar to the source spectrum. The selection is based on the atmospheric parameters of the source which are assumed to be known from the photometry.

* Compute the correlation peak of the source and template spectra. This is performed by shifting step by step the template spectrum in radial velocity. At each step:
  - the template spectrum is interpolated at the wavelength of the source spectrum.
  - the cross-correlation coefficient of the template (shifted at a given radial velocity) and source spectra is computed.

Note: to save computing time the correlation peak is computed
3.3 Radial velocity cross-correlation (3-03-S)

iteratively: in two steps. First with a large step of 10 km/s and then with a step of 2 km/s around the first estimate of the source radial velocity.

* The radial velocity of the source is given by the location of the maximum of the correlation peak. It is determined by fitting the top of the correlation peak with a 2nd order polynomial.

* The radial velocity of the source is corrected for the zero point.

4- Implementation
-----------------

The method deriving the radial velocity is:

```java
void deriveRadialVelocity(Spectrum sourceSpectrum, AtmosphericParameters atmoParameters, double zeroPointCorrection)
```

It has been assumed to be a method of the "RVSResult" class (of the GDAAS data model). Therefore it return no value, but modify/define the "radial velocity" (and in a later version, the error on the radial velocity) attribute of an object of "RVSResult" type.

It has three input parameters: (i) the source spectrum, (ii) the atmospheric parameters of the source and the (iii) radial velocity zero point correction. The list of templates and the template spectrum have been assumed to be auxiliary data and are queried inside the "deriveRadialVelocity" method. The absolute path to the auxiliary data is defined at the beginning of the "deriveRadialVelocity" method.

For the purpose of the development of the algorithm, a "Test" class has been written. It contains the "main" method and read/define the input parameters of the "deriveRadialVelocity" method. The only function of this class is to mimic the GDAAS environment and it should not be included into GDAAS.

The "deriveRadialVelocity" method uses objects and methods of 4 main classes:
* Spectrum.java
  - The attributes describe a spectrum: length / table of wavelength / table of flux.
  - The methods performed the following tasks:
    - void readSpectrum(String fileName) : read a spectrum file
    - void copySpectrum(Spectrum spectrumToCopy) : copy a spectrum
    - void rvShiftSpectrum(double radialVelocity) : shift a spectrum in wavelength (according to radial velocity)
    - void interpolateSpectrum(Spectrum referenceSpectrum) : interpolate a spectrum at the wavelength of another spectrum
3.3 Radial velocity cross-correlation (3-03-S)

* AtmosphericParameters.java
  - The attributes of this class are the atmospheric parameters of the source: effective temperature, surface gravity, metallicity
  Note: this class should be (is already?) merged with a wider class of the GDAAS data model.

* AuxiliarySpectraLibrary.java
  - The methods of this class performed two tasks:
    - void readLibraryParameters() :
      query the list of auxiliary/template spectra and their atmospheric parameters.
    - String selectAuxiliarySpectrum(AtmosphericParameters atmoParams) :
      select the auxiliary/template spectrum whose parameters are the most similar to those of the source.

* CorrelationPeak.java
  - The correlation peak is described by:
    its number of points (size) / the radial velocity shift of the template and the value of the cross-correlation in each point / the index of the estimated position of its maximum.
  - The methods that are applied to the correlation peak are:
    - double computeCorrelationCoefficient(Spectrum sourceSpectrum, Spectrum auxiliary Spectrum) : compute the cross-correlation coefficient of the source and template spectra.
    - void determineMaximumLocation() : determine the index of the first estimate of the position the maximum of the correlation peak.
    - double radialVelocityByPolynomialFit(int fitHalfWidth) : determine the radial velocity by fitting with a 2nd order polynomial the top of the correlation coefficient.
    - void resizeCorrelationPeak(int extraHalfNumberOfSteps, double newStepSize) : resize and allocate free space to the correlation peak to allow to refine its the calculation with a finer radial velocity step around the first estimate of the position of its maximum.

Two other classes have been written to contain physical constants and mathematical method:
* Constants.java : LIGHTVELOCITY
* MathLib.java : linear interpolation
Those two classes should of course be replaced by their GDAAS counterparts.

5- Spectra libraries
--------------------
Two libraries of spectra are provided:
* 49 source spectra to be used by GASS to simulate RVS observations.
They correspond to 7 stellar types:

<table>
<thead>
<tr>
<th>Teff</th>
<th>log g</th>
<th>[Fe/H]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4000</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>5000</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>6000</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>7500</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>10000</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>15000</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>20000</td>
<td>4.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

and 7 magnitudes: $V = 8, 10, 12, 14, 16, 17$ and $18$.

* 1237 synthetic template/auxiliary spectra.

A spectrum file (source and template/auxiliary) started with a header listing:
- the star effective temperature
- the star surface gravity
- the star metallicity
- the star radial velocity
- the star magnitude (not relevant for the template/auxiliary spectra)
- the length of the spectrum
- the width of the spectrum (in the present case, source and auxiliary spectra are all 1 pixel wide)

Then follow as many wavelengths as the length of the spectrum, followed by the same number of fluxes.
4 Minor planets

4.1 Minor planets motion detection (4-01-C)

4.1.1 Function

To detect and estimate any (linear) motion of a detected source over the Astro field transit.

4.1.2 Method

Regression analysis of AF1–11 centroid positions with respect to the current (possibly raw) attitude.

4.1.3 Input data

- elementary data (centroid positions) for a single transit over AF1–11
- attitude data for the corresponding time interval
- calibration data for AF and the relevant time interval

4.1.4 Output data

- flag whether motion was detected or not
- estimated inertial rate of motion, including uncertainty

4.1.5 Data access requirements

Elementary data:
  READ batch / time and source ID / AF data for one transit
  WRITE none

Sources:
  READ single / ID / not matched to a previous source
  WRITE single / same as read / update

Attitude:
  READ batch / time
  WRITE none

Calibration:
  READ batch / time and CCDs (AF1–11)
  WRITE none
4.1.6 Remarks

1. For detecting source motion a number of approximations can be made, thus avoiding detailed calculation of proper directions. A transit across AF takes 36 s, during which the proper motion of a star is at most 12 $\mu$as, while the change in stellar aberration is less than 150 $\mu$as (0.0034 pixel). Both effects can be neglected for the present algorithm. Thus only the changes caused by attitude and geometrical calibration need to be considered (chromaticity can be neglected, TBC).

2. In a later stage, the algorithm may be improved by using also ASM and BBP data.

4.1.7 Provider

Mignard

4.1.8 Schedule

Expected algorithm delivery: July 2003
4.2 Minor planets cross-identification (4-02-S)

4.2.1 Function

To identify minor-planet observations that belong to the same object.

4.2.2 Method

TBD

4.2.3 Input data

TBD

4.2.4 Output data

TBD

4.2.5 Data access requirements

TBD

4.2.6 Remarks

1. Cross-matching over a limited time span (e.g., between a few successive field transits) can probably be made simply on the basis of observed motions (extrapolation), combined with photometric data. Over longer time spans, cross-matching requires some kind of orbital fit, in which case the task overlaps with 4-03-S (Minor planets orbital solution).

4.2.7 Provider

Mignard

4.2.8 Schedule

Expected algorithm delivery: September 2004
4.3 Minor planets orbital solution (4-03-S)

4.3.1 Function

To fit an orbit to (matched) observations of a minor planet.

4.3.2 Method

Robust least-squares fitting of orbital elements to observed directions (and angular rates, TBC) for the transits of a single source, classified as a minor planet.

4.3.3 Input data

- source parameters
- elementary data for the source (including estimated angular rates), possibly collected per transit
- attitude
- geometrical calibrations
- satellite ephemeris
- solar system ephemerides (to calculate perturbations)

4.3.4 Output data

- osculating orbital elements (or barycentric position and velocity) at the reference epoch

4.3.5 Data access requirements

Elementary data:
  READ batch / link from source / AF transits
  WRITE none

Sources:
  READ single / ID / minor planet
  WRITE none

Attitude:
  READ batch / link from source
  WRITE none
4.3 Minor planets orbital solution (4-03-S)

Calibration:
   READ batch / geometric AF
   WRITE none

Global parameters:
   READ all
   WRITE none

Ephemerides:
   READ batch / link from source
   WRITE none

Minor planets orbital elements:
   READ none or single
   WRITE single / same as read / insert or update

4.3.6 Remarks

1. This algorithm overlaps with 4-02-S (Minor planets cross-identification), since the cross-
   identification of well separated observations can only be made on the basis of trial orbit
determinations.

4.3.7 Provider

Mignard

4.3.8 Schedule

Expected algorithm delivery: September 2004
Table 1: Discrete source classes. These classes are assigned mutually exclusive and normalised probabilities by the Discrete Source Classifier (DSC).

<table>
<thead>
<tr>
<th>Class No.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>single star</td>
</tr>
<tr>
<td>2</td>
<td>single QSO</td>
</tr>
<tr>
<td>3</td>
<td>single galaxy</td>
</tr>
<tr>
<td>4</td>
<td>single solar system object (a rock)</td>
</tr>
<tr>
<td>5</td>
<td>composite stellar system (physical or optical or both)</td>
</tr>
<tr>
<td>6</td>
<td>composite non-stellar system (i.e., at least one component is not a star)</td>
</tr>
<tr>
<td>7</td>
<td>other</td>
</tr>
</tbody>
</table>

5 Classification and Parametrization

Classification refers to the process of assigning a discrete source class to an observed source, or more generally to assigning the probabilities of different classes to the source. Table 1 describes the discrete sources considered here (from ICAP-CBJ-011, version 1).

Parametrization refers to the process of estimating the parameters (usually continuous quantities) relevant for describing a source, including the astrophysical parameters (AP). The following parameters are envisaged:

- probability of being a composite source\(^6\)
- astrometric parameters (6) and errors
- probability of being photometrically variable across all data collected so far; one measure for each distinct filter
- for each class
  - class probability
  - APs and associated errors. In general, this will consist of more than one set of APs, e.g., if the source is composite (classes 5 and 6 in Table 1) then there will be a set of APs for each component. However, even non-composite sources may give multiple APs for a single class (see below)
- links to observations, distinguishing between
  - where source is spatially single
  - where source is multiple
- history of tasks executed on this object; these are automatically updated whenever this object is written to.

For every source, APs may be determined for more than one class.

\(^6\) We could potentially get a conflict between this parameter and the source class probabilities.
Note: Much of the definitions in Section 5 are directly taken or derived from ICAP-CBJ-011 (draft 28.11.02). The term ‘window object’ used in that document was in the present document changed to ‘observation’, ‘transit’ or ‘elementary data’, as the case may be, for improved consistency with other definitions.
5.1 Ingestion level classifier, ILC (5-01-C)

5.1.1 Function

To determine the discrete class (Table 1) of a source from a single observation (or transit), and a more detailed classification (e.g. spectral type for stars), using data from a single instrument (MBP, BBP or RVS).

5.1.2 Method

Pattern recognition, with appropriate attention paid to non-count inputs.

5.1.3 Input data

- BBP, MBP or RVS photometric data from a single transit
- data from the onboard detection (OBD) algorithm (in satellite data stream) giving information on:
  - whether or not the source is multiple (from the star mapper)
  - whether or not the source is extended (probability that it is not a point source)
  - whether or not the source is a high proper motion source
  - any preliminary data reductions from the star mappers/OBD

5.1.4 Output data

- class probabilities (for the classes defined in Table 1 — written to elementary data
- more specific classifications (e.g., spectral type for stars) — written to elementary data

5.1.5 Data access requirements

Elementary data:
  READ batch / time and source / ASM + (BBP, MBP or RVS photometry)
  WRITE batch / those read / update

5.1.6 Remarks

1. This algorithm exists primarily to serve other algorithms, in particular object matching and photometric analysis. Therefore, the information which it provides is driven primarily by these procedural requirements rather than astrophysical requirements. Input is required from other algorithm developers – in particular object matching and photometric calibration – to determine the requirements of this algorithm.
2. This algorithm will probably first pipe data to the DSC (or a simplified version thereof) to determine the class (in the sense of Table 1) before then determining more detailed classifications specific to each class.

3. In practice we may have separate algorithms for each of BBP, MBP, RVS.

4. It still has to be established how this algorithm would use information from the OBD on multiplicity and extended objects.

5. It is unclear whether this algorithm is really needed. Its tasks can be fulfilled by DSC and SSP, provided that both of these are core algorithms (CB-J, 24 Feb 2003). At GST-7 it was agreed not to implement the ILC in GDAAS-2, because no classification is required for the object matching algorithm 1-01-C.

5.1.7 Provider

Bailer-Jones

5.1.8 Schedule

Not implemented in GDAAS-2 (see Remark 5).
5.2 Multiple source detector, MSD (5-02-C)

5.2.1 Function

To determine, for a given source object, which of its observations contain other sources. Decides whether classification and parametrization for this source will use (1) non-multiple window objects, (2) multiple window objects, or (3) both.

5.2.2 Method

Simply interrogates the window objects of the present source to see what other sources are present in the window objects. This is done by means of the links contained within each window object which point to all sources contained within it. This information will have been set by the OBD and the initial object matching algorithm. At the simplest level, the interrogation would consist of just deciding whether any windows contain additional sources, but at a more advanced level it would also use the magnitude difference to decide to what extent the additional sources contaminate. Further PSF fitting to find additional sources may not be necessary.

5.2.3 Input data

- window objects for a given source, coming from one or more scans from AF/BBP (or AFSM), MBPSM and/or RVSSM (via link to window objects in the source object)

5.2.4 Output data

- one list for each instrument (AF/BBP, MBP, RVS) listing which scans (i.e., which window objects) are to be used for multi-epoch classification (i.e., by DSC). Written to source object. If this list consists of window objects which have been found to be multiple (i.e., contain other sources, not resolved in these window objects) then this fact is also output and recorded in the source object under the source object parameter ‘probability of being a composite source’

5.2.5 Data access requirements

Elementary data:
  READ batch / link from source
  WRITE none

Sources:
  READ single / ID
  WRITE single / same as read / update
5.2.6 Remarks

1. The function of this algorithm may be partially performed by other algorithms considered for inclusion in the final GAIA data processing system. It may also be more appropriate to remove it from ‘classification’ to ‘astrometry’.\textsuperscript{7}

2. This algorithm makes no use of astrometric data. Use of astrometric data to detect \textit{composite sources} (rather than \textit{multiple sources}) is undertaken by DSC, using the scan lists generated here.

5.2.7 Provider

Bailer-Jones

5.2.8 Schedule

Expected algorithm delivery:
April 2004 (preliminary)
September 2003 (update)

\textsuperscript{7} CB-J: It is not clear to me whether the different sample sizes in the along scan and across scan directions was considered by the object matching routine in GDAAS-1, and thus whether this problem has yet been addressed. It is not mentioned under 1-01-C (Improved object matching), so I assume not.


5.3 Discrete source classifier, DSC (5-03-C)

5.3.1 Function

To determine the class (defined in Table 1) of a source based on single or multi-epoch data from one or more instruments (BBP, MBP, RVS).

5.3.2 Method

The basic algorithm will be pattern matching to determine discrete classes, but with particular attention paid to incorporation of astrometric data.

5.3.3 Input data

- photometric data from one or more scans from MBP, BBP and/or RVS, from window objects (accessed via links from source object using lists created by MSD, if present, otherwise all window objects are used)
- any existing class probabilities in the database, coming from previous executions of the present algorithm (from source object) or from other algorithms or databases (e.g., SDSS for QSOs) or from any GAIA onboard input catalogue (a limited catalogue may be included to ensure GAIA windows certain objects of special interest or relevance)
- probability that source is a multiple or composite source (from source object)
- astrometric data (if present), specifically parallax, proper motions and any indications from binary star astrometry algorithms that source is an unresolved binary (from source object)
- (TBC) interstellar extinction obtained from other catalogues or a Galaxy model, in which case source position must also be read from the source object.

5.3.4 Output data

- mutually exclusive and normalised probabilities that the source corresponds to each of the classes defined in Table 1 — written to source object
- details of data used to achieve classification — written to source object

5.3.5 Data access requirements

Elementary data:
READ batch / link from source
WRITE none
Sources:
   READ single / ID
   WRITE single / same as read / update

Auxiliary data (galaxy model):
   READ single / position
   WRITE none

5.3.6 Remarks

1. In practice we may have multiple algorithms depending on which data are available or selected from the database on a given source (i.e. which classification mode in the parlance of ICAP-CBJ-007). However, we may not want to use RVS data as an input at all.

2. Temporal variability will not be used: a robust averaging of multiple scans from a given instrument is performed instead.

5.3.7 Provider

Bailer-Jones/Christlieb

5.3.8 Schedule

Expected algorithm delivery:
July 2003 (preliminary)
September 2004 (update)
5.4 Single Star Parametrizer, SSP (5-04-S)

5.4.1 Function

Given a list of sources assumed to be single stars with given MBP photometry, determine for each the following astrophysical parameters (APs): $T_{\text{eff}}$ (effective temperature); $\log g$ (surface gravity); [Fe/H] (metallicity); $A_v$ (extinction).

5.4.2 Method

Simple minimum distance method. APs are assigned to each program star based on the APs of a set of template stars. (A program star is an unclassified GAIA observation; a template star is one with existing classifications.) The assigned AP (for each of the four APs to be determined) is the mean of the APs of all templates found within unit search radius in the data (photometry) space, where each data dimension is scaled by the 1-$\sigma$ photometric error for that program star. Noise free templates are used. Note that a minimum and a maximum number of nearest neighbours which can be used in the mean is defined, currently set at 1 and 10 respectively.

5.4.3 Input data

To classify a star, the following must be provided:

1. MBP area normalised photoelectron counts, and
2. the corresponding additive errors.

Each of the above is N-dimensional vector where N is the number of filters. The data are area normalised photoelectrons. This means that the photoelectron counts sum to unity, i.e. the original number of photoelectrons obtained in each filter has been divided by the total in all filters.

The classification code works in batch mode, i.e. on a set of stars to be classified. Three input files are required:

1. template data file, listing the area normalised photoelectron count data and APs of the templates,
2. program data file, listing the area normalised photoelectron count data on the stars to be classified,
3. program error file, listing the additive errors in the program data.

In all cases, the files have a 5-line header followed by one line per star. These lines have 1 ID (identification) field (a unique string), Y output fields and X input fields (reals). For the specific file provided, X=11 (no. filters) and Y=5 (G magnitude + 4 APs). These values must agree with those specified in the input specfile (see below). (The G magnitude
is not, of course, an “output” from the classification model, but it is retained in the input file for convenience and ignored by the classification program.) Note that the program file may contain the “true” APs of the program stars (e.g. for synthetic data). This is the case for the test data supplied. In reality, these would be missing. The order of the APs is: $A_v$, $[\text{Fe/H}]$, $\log g$, $T_{\text{eff}}$.

Various checks are made by the code to ensure that the input data are well behaved, in particular that the data are positive.

The running of MDM is controlled by a specfile. This lists the input files to use plus parameter settings for nearest neighbour searching, data scaling etc. Details of what these specifications do are given in the file `mdm_manual`. It also gives a summary of the code subroutines/operation. Input is performed by: `dataread.c`, `openfile.c`, `specread.c` and `mdm.c`.

### 5.4.4 Output data

A screen dump plus two output files are produced.

For each star, the four APs are determined. These are written to a file with suffix `.mdm`, which has an 11 line header and one line for each star in the program data file. The remainder of the file has 13 columns: 1 ID plus 3 columns per AP. For each AP, the following are listed: value of AP determined by MDM; the true value of the AP (as listed in the program data file); the difference between these two numbers.

The `.knn` file lists other diagnostic information which can be ignored.

This version of MDM does not provide uncertainties in the APs.

All output is done by `mdm.c`.

### 5.4.5 Data access requirements

These are described above in the sections on input and output data. (I am unsure how these correspond to “READ batch” and “WRITE batch” statements.)

### 5.4.6 Remarks

**Source code, compilation and execution**

The code, MDM version 1.03, is written in C and compiled with gcc under the control of a Makefile:

```
make mdm
```

To execute:

```
mdm SSP_v01.spec
```
5.4 Single Star Parametrizer, SSP (5-04-S)

where **ssp_v01.spec** is the specfile file. MDM does its classifications in real time based only on the three supplied inputs files and the specification file. There is no additional file of model parameters (unlike, for example, a pre-trained classifier).

**Assumed instrument model**

The instrument set-up assumed by the current specfile is summarized below. It is largely determined by the test data supplied, which in this case is the ICAP blind testing cycle 1 data (which in turn used version 1 of photsim and the GAIA-1 instrument model). For more details of photsim see ICAP-CBJ-004 and for more information on the cycle 1 blind testing data see ICAP-CBJ-010, both available from [http://www.mpia.de/GAIA/](http://www.mpia.de/GAIA/) under ICAP documents. Also see Remarks below.

- Spectro mirror area = 1.19m²; wavelength independent instrument transmission = 0.9
- CCD1B for filters with $\lambda_{\text{central}} < 5500\text{Å}$; CCD2 for $\lambda_{\text{central}} \geq 5500\text{Å}$
- 1X filter system (which has 11 filters)
- mean of 100 focal plane passages (epochs); 45s integration time for all filters per passge. (This only affects the noise level in the program data: it does not effect the template data if that is noise free.)
- noise sources: Poisson from source and sky; CCD read out noise; margin factor 1.2

**Testing data**

Three data files (corresponding to the three input files described above) are provided for testing purposes. They contain single stars over a range of the four APs with added noise to simulate G=20. These data are exactly the data used in cycle 1 of the blind testing of filter systems (but with the photoelectron counts area normalised).

- blindtest1-stars-a_1x_01_g=20_.nphot.in  template data file (noise free)
- blindtest1-stars-b_1x_01_g=20_1.nphot.in  program data file (noisy)
- blindtest1-stars-b_1x_01_g=20_1.nerr.in  program error file

**General remarks**

1. It is implicitly assumed (in terms of GDAAS data dependencies) that the DSC (“Discrete Source Classifier”) has been run to draw up a list of single stars to serve as the input to SSP. DSC in turn implicitly assumes some kind of photometric reduction has been done to bring the photometry to a common time-independent system. Of course, the present SSP code provided will run independently using the provided test data. If data are provided to SSP which do not correspond to a single star, SSP will nonetheless assign it stellar APs based on its photoncount data, whatever that may be.

2. The classification algorithm used is very simple and far too simple for the real GAIA problem. Thus the results will be poor, so this should only be considered as a placeholder algorithm. This (first version) of the classifier is supplied to give some indication of how
the classifier code could operate. In particular, it makes use of a real time search on a template of objects (albeit the most inefficient one, namely an exhaustive search).

3. The fact that the test data assume an out-of-date instrument model (GAIA-1 rather than GAIA-2) is essentially irrelevant for this first introduction of SSP into GDAAS. This is partly because it is assumed that this version of SSP will only ever be run on the test data supplied (this was the impression obtained from the GDAAS Algorithms meeting held in Barcelona in April 2003). However, it should be noted that the classification model is quite general: the instrument model is reflected only in the data and not in the code.

4. The code was prepared taking into account the C coding guidelines given in GMV-GDAAS2-SCG-003 version 1.0 dated 11.04.03, which was the most up-to-date version available on 02.06.03. As the code is based on existing code, the guideline of not making unnecessary changes to avoid introducing coding errors was taken advantage of.

5.4.7 Provider

Coryn A.L. Bailer-Jones (calj@mpia.de).

5.4.8 Schedule

First version of code to be delivered by 1. July 2003. An updated version (with increased functionality, performance etc.) will be supplied by September 2004.
5.5 Single quasar parametrizer, SQP (5-05-S)

5.5.1 Function

To determine astrophysical parameters for a single QSO, i.e., a source of class 2 (Table 1), using single or multi epoch BBP, MBP and/or RVS data. (Preliminary description, CBJ 06.05.03.)

5.5.2 Method

A pattern matching method, such as an artificial neural network (ANN) or a minimum distance method (MDM).

5.5.3 Input data

- calibrated photometric data plus error from one or more scans from (1) BBP, (2) MBP, or (3) BBP+MBP. If more than one scan is available, this is averaged (unless a running sum/average is stored in the database).
- (TBC) initial estimates of interstellar extinction from other databases or a Galaxy model, in which case source position must also be read from the source object.

5.5.4 Output data

- a set of APs (and associated error estimates), in particular the redshift, z, and QSO specific parameters: alpha (the slope of the continuum); beta (the line strength).
- details of data used to achieve the parametrization

5.5.5 Data access requirements

Elementary data:
  READ batch / link from source
  WRITE none

Sources:
  READ single / ID
  WRITE single / same as read / update

Auxiliary data (galaxy model):
  READ single / position
  WRITE none
5.5.6 Remarks

Details of this algorithm purposely remain vague, awaiting experience gained from the implementation of the SSP and DSC classification algorithms. In principle, SQP will behave in a very similar way to SSP, just on single quasars rather than single stars. The inputs from the Galaxy model are optional. They would be a useful way of testing the notion of linking GDAAS data processing to an external model.

5.5.7 Provider

Bailer-Jones (co-ordination)
Smette/Claeskens (authorship)

5.5.8 Schedule

Expected algorithm delivery:
April 2004 (preliminary)
Appendix A: Overview of Astro calibration parameters

This Appendix describes the proposed calibration model in GDAAS-2 for the Astro instrument. The description pertains to algorithms 1-05-C (Improved instrument modelling), 1-06-C (Detailed geometrical calibration), 1-07-C (Chromaticity calibration), and 2-01-C (Photometric calibration).

The need to define a single geometric origin and a single photometric system for the whole mission makes it necessary to do the instrument calibration globally. That is, even if the calibration is assumed to vary on a certain time scale (say, a few months), so that a separate set of parameters is determined for each such time interval, all the time intervals must ultimately be tied together to represent a constant system. At least part of the instrument calibration must therefore be made as a global updating, rather than a local one, and thus executed as part of the GIS.

A.1 Proposed calibration units, temporal and spatial scales

Recall that in GDAAS-1 the modelling of the Astro instruments included just four parameters per ‘calibration unit’ (CCD × time interval), viz.: $\Delta \eta$ and $\Delta \zeta$ for the geometric position of a CCD, $\mu$ for the photometric zero point of a CCD, and $\Gamma$ for the chromatic dependence of $\eta$. The time interval considered in a calibration unit was one month (TBC). Because of the physical separation of focal planes in GAIA-1, it was appropriate to assume that a separate calibration was needed for Astro-1 and Astro-2.

For GDAAS-2 it is the aim to have a realistic instrument model in terms of complexity and the number of parameters to be determined. Such a model must probably assume that each pixel column (1966 per CCD) needs to be calibrated both geometrically and photometrically (‘column’ here means a single along-scan line of pixels). This is still possible on a time scale of one or a few months if stars down to at least 15th magnitude can be used. Furthermore, the photometric calibration should take into account colour terms, i.e. that the response functions (QE times transmittance) are not identical for all columns/CCDs in a given row. Finally, a more realistic chromaticity calibration needs to take into account all five BBP filter bands, rather than a single colour index.

It is also assumed that GDAAS-2 shall include the full treatment of sample data (not just the elementary observations as in GDAAS-1), including the calibration of the relevant line spread functions (LSF). Analysis of individual samples is required in algorithms such as the double star analysis.

The new Astro design with superposed fields also has a major impact on how the instrument model is formulated. Clearly most of the column-to-column variation in geometry and sensitivity must originate in the CCD chip itself rather than the optics, and therefore can be expected to be the same in both fields. This effectively halves the number of such

---

8 The mean rate of stars with $G \leq 15$ hitting a given CCD column, 30 $\mu$m wide, is about 60 per day.

9 However, the calibration of point spread functions (PSF) is not included — the cross-scan distribution of light is assumed to be fixed and known in GDAAS-2 (TBC).
A.1 Proposed calibration units, temporal and spatial scales

Table 2: Summary of Astro calibration parameters. Small-scale parameters (index \( \kappa \)) are defined per pixel column and are the same in both fields for common CCDs; large-scale parameters (index \( k \)) are defined per CCD but separately in Astro-1 and Astro-2. Short, medium and long time scales are of order 1 day, 3 months, and 5 years.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Applicable to</th>
<th>Spatial scale</th>
<th>Temporal scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>small</td>
<td>large</td>
</tr>
<tr>
<td>Geometry:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>along-scan CCD pos. ( \Delta \eta )</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>along-scan column pos. ( \delta \eta )</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>across-scan CCD pos. ( \Delta \zeta )</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Chromaticity:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>along-scan CCD pos. ( \Gamma )</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Response functions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>point-spread function ( P )</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>line-spread function ( L )</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Photometry:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCD sensitivity ( A )</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>column sensitivity ( a )</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

† the calibration of PSF is not considered in GDAAS-2 (TBC)

calibration parameters, which is of course beneficial both for the data processing and the accuracy. On the other hand, the optics will also produce some large-scale geometric and photometric distortion, which may be different in Astro-1 and Astro-2. Thus the instrument model should consist of two parts, one (small-scale) part common to both fields, and a (large-scale) part which is different in the two fields. Since both parts in principle describe the same kind of dependencies, the separation between them must be made unique by suitable constraints. A natural way to achieve this is to make the large-scale and small-scale representations orthogonal to each other. Each orthogonality condition imposes a linear constraint on the parameters. For instance, if the large-scale parameter is simply a constant for a given calibration unit, then the orthogonality condition implies that the average small-scale parameter for that calibration unit must be zero.

A reasonable model for GDAAS-2 is to assume that the effects on pixel-column level are common to the Astro fields (they represent some \( 180 \times 1966 \approx 354 \text{,}000 \) points, each with several calibration parameters), while those on CCD level need to be separately calibrated for Astro-1 and Astro-2 (i.e., some \( 2 \times 170 = 340 \) points). Geometric and photometric calibrations are needed on both scales, while chromaticity and PSF/LSF calibration are only needed on the large scale (TBC).

The definition of ‘calibration unit’ also needs to be revised. Recall that a single index \( k \)
was used to identify a certain part of the instrument (say, a CCD) in a given time interval. With the division of calibration parameters into large-scale and small-scale parameters, it is not obvious that the time intervals considered for the calibration units must remain the same. Indeed, it is more natural to assume that the large-scale parameters must (and can) be calibrated on a shorter time scale than the small-scale parameters. On the other hand, some large-scale properties (across-scan geometry, chromaticity, PSF/LSF) probably need a longer time scale for their calibration, or can perhaps be assumed constant over the mission (across-scan geometry). Thus, the separation according to time scales may be different from the one based on spatial scales.

It is proposed that the calibration is essentially made on three different time scales: short ($\sim 1$ day), medium ($\sim 3$ months), and long ($\sim 5$ years). Table 2 is an overview of the various Astro calibration parameters and their spatial/temporal scales. We shall continue to use index $k$ to identify the large-scale calibration units (on whatever time scale), and introduce index $\kappa$ to identify the small-scale calibration units.

### A.2 Use of a standard (instrument) photometric system

An important issue for a more realistic instrument model concerns the definition of a standard (instrument) photometric system. In GDAAS-1 the spatial and temporal variations of the photometric response functions were not considered, and consequently the definition of the magnitude scales (call them $G$, $B_1$, $B_2$, $B_3$, $B_4$, $B_5$ for the AF and BBP bands) was trivial, requiring only a single magnitude zero point ($\mu$) per calibration unit; this was fixed by simple averaging over a subset of photometric standard stars. In GDAAS-2 the variations of the response functions are part of the model, and there is then the question which particular response function (or actually calibration unit) should be regarded as the ‘standard’ one, to which other units are to be transformed.\(^{10}\) The obvious answer seems to be that the standard system should not be based on any particular calibration unit, but on the mean properties of all the relevant units over the whole mission (or at least a large part of the mission, say the first few years).

### A.3 Geometric and chromatic calibrations

The first four calibration items in Table 2 are concerned with modelling the centroid position (along and across scan) as function of time, position in field and spectral composition of the source. This presupposes that there is a unique definition of the centroid, which in turn is linked to the definitions of the LSF and PSF. These are discussed below in

\(^{10}\) It can be argued that the present approach, i.e. defining a standard GAIA photometric system to which the observations are transformed, is not an optimal way to use the data. Any transformation will be inexact and introduce biases, especially for objects with peculiar spectra. Indeed, for some purposes like the astrophysical parameterisation, it might be better to model the observed fluxes directly in terms of model spectra and the varying response functions. For such reasons it is important that the ‘raw’ fluxes are preserved in the database. However, for any external user, a well-defined standard system is indispensable and it also makes e.g. the detection and analysis of variable stars much easier. Furthermore, the definition of a standard system appears to be a necessary ingredient of the instrument calibration, as described here.
A.3 Geometric and chromatic calibrations

In general, the following indexing conventions are used: \( i \) denotes the source, \( l \) the observation (representing a single CCD crossing, or a window), \( n \) the individual sample within a window, \( k \) and \( \kappa \) the large-scale and small-scale calibration units, and \( m \) the photometric filter. When relevant, *-to-one relations are indicated by functions, as in \( i(l) \) = the source for observation \( l \).

A.3.1 Basic observation equation

Let \( \eta_0^k \) be the nominal (reference) along-scan instrument angle of the unit \( k = k(l) \) associated with observation \( l \), and \( \kappa(l) \) the corresponding mapping from \( l \) to \( \kappa \). The proposed model for the along-scan geometric and chromatic calibration is

\[
\eta_{\text{obs}}(l) = \eta_0^k(l) + \Delta \eta_k(l) + \delta \eta_{\kappa}(l) + \sum_{m=1}^{M} \Gamma_{k(l)m} \phi_{i(l)m} + \text{(noise)}
\]  

(2)

where \( \Delta \eta_k \) is the large-scale parameter, \( \delta \eta_{\kappa} \) the small-scale parameter, \( \Gamma_{km} \) \((m = 1 \ldots M)\) the chromaticity parameters for the \( M \) BBP filters, and \( \phi_{im} \) the normalised standard fluxes (such that \( \phi_{i1} + \phi_{i2} + \cdots + \phi_{iM} = 1 \)).

The attitude is implicitly defined by the nominal instrument angles, so the along-scan instrument angle \( \eta_{\text{calc}} \) calculated from the observed CCD transit time \( t_l \) and current source parameters, attitude, and global parameters [see Eq. (11) in GAIA–LL–034] should (apart from noise in the latter parameters) equal \( \eta_0^k \). With \( H_l \equiv \eta_{\text{calc}} - \eta_0^k(l) \) we therefore have the general along-scan observation equation\(^\text{11}\)

\[
\Delta \eta_k(l) + \delta \eta_{\kappa}(l) + \sum_{m=1}^{M} \Gamma_{k(l)m} \phi_{i(l)m} \sim H_l.
\]  

(3)

In this equation the normalised BBP fluxes \( \phi_{im} \) are regarded as known quantities for any transit across the AF or BBP. The unknowns are \( \Delta \eta_k, \delta \eta_{\kappa} \) and \( \Gamma_{km} \).

The along-scan geometric and chromatic calibration is made by iteratively solving the complete set of these observation equations in a robust, weighted least-squares sense, taking into account the orthogonality constraints discussed below.

A.3.2 Orthogonality constraints

The three deterministic terms on the right-hand side of (3) are not uniquely separated without additional constraints. For instance, because \( \sum_m \phi_{im} = 1 \) it is clear that if

\(^\text{11}\) An observation equation (or equation of condition) is here consistently written in the form \( f(x|y) \sim f_{\text{obs}} \), where \( f \) is a known function of the unknown parameters \( x \) and any other given data \( y \), \( f_{\text{obs}} \) the observed value of the function, and ‘\( \sim \)’ means ‘equals, apart from the noise’. Note that whatever is placed on the right-hand side of the ‘\( \sim \)’ is regarded as given (by observation). In least-squares estimation, a set of observation equations are processed to give an estimate of \( x \) and its covariance matrix.
A.3 Geometric and chromatic calibrations

the $M$ chromaticity parameters $\Gamma_{km}$ are increased by the arbitrary amount $c$ and $\Delta \eta_k$ simultaneously decreased by $c$, then (3) remains valid. The separation between $\Delta \eta_k$ and $\Gamma_{km}$ can be made unique by requiring that

$$\sum_m \Gamma_{km} = 0 \quad \text{for all } k.$$  

(4)

At the same time it is clear that adding this constraint to the least-squares problem does not in any way ‘force’ the solution. We write the constraint more concisely as

$$\langle \Gamma_{km} \rangle_m = 0 \quad \text{for all } k,$$  

(5)

where $\langle \alpha \rangle_\beta$ means the (unweighted) average of quantity $\alpha$ over the possible range of index $\beta$.

Similarly, the separation between the large-scale parameters $\Delta \eta_k$ and the small-scale parameters $\delta \eta_k$ is made unique by means of the constraint

$$\langle \delta \eta_k \rangle_{\kappa \in K_k} = 0 \quad \text{for all } k,$$  

(6)

where $K_k$ is the set of $\kappa$ values for which it is possible to have an observation $l$ with $k(l) = k$ and $\kappa(l) = \kappa$.

Finally the origin of the $\eta$ coordinate is defined by the constraint

$$\langle \Delta \eta_k \rangle_k = 0.$$  

(7)

A.3.3 Iterative solution of observation equations

It is now clear that the three kinds of along-scan calibration parameters can be solved iteratively by processing the following modified versions of (3):

$$\Delta \eta_k \sim H_l - \delta \kappa - \sum_m \Gamma_{km} \phi_{km}$$  

(8)

$$\delta \eta_k \sim H_l - \Delta \eta_k - \sum_m \Gamma_{km} \phi_{km}$$  

(9)

$$\sum_m \Gamma_{km} \phi_{km} \sim H_l - \Delta \eta_k - \delta \kappa$$  

(10)

where, additionally, in each iteration, the parameters are re-normalised by means of the orthogonality constraints (5)–(7). The normal equations for each kind of calibration parameter can be accumulated in parallel, since the main effort is to calculate $H_l$ for each observation in turn.

A.3.4 Across-scan geometric calibration

This could remain as in GDAAS-1, i.e. a single parameter $\Delta \zeta_k$ per large-scale unit. It is only relevant for the ASM, where the across-scan coordinate is actually measured (TBC). The origin of the $\zeta$ coordinate is fixed by the constraint

$$\langle \Delta \zeta_k \rangle_k = 0.$$  

(11)
The observation equations,

\[ \Delta \zeta_k \sim \zeta_l^{(\text{calc})} - \zeta_l^{(\text{obs})} \]  

(12)
can be processed in parallel with the corresponding equations for the along-scan calibration.

### A.4 Response function (LSF/PSF) calibration

The geometric (and chromatic) calibrations described above refer to the centroid of the detected image. Conversely, the LSF and PSF are defined relative to a positional origin which must correspond to this centroid. Unfortunately, the precise definition of centroid depends on the location estimation method used and for optimal estimators such as the Maximum Likelihood one described in GAIA–LL–032, the result may in principle differ e.g. between bright and faint stars, or between different levels of background noise. From the viewpoint of instrument calibration this is clearly not a very satisfactory situation, as it could for example result in magnitude-dependent geometric calibrations even for perfectly linear detectors, if a slightly erroneous response function is fitted. It is not clear how this should best be solved. Here, the following solution is adopted.

Firstly, any LSF or PSF must be geometrically normalised, just as it is photometrically normalised to unit volume or area. The geometric normalisation means that the origin is defined by some simple mathematical prescription. A possible prescription could be to make the first moment zero (origin = centre of gravity), but this is too sensitive to the ill-defined outer parts of the function. The median is also problematic, because of truncation. The mid-point at half maximum does not suffer from truncation but is too local and therefore sensitive to interpolation errors. Here, we will use the ‘minimum difference mirror point’.\(^\text{12}\) Secondly, the centroid of the observed sample data is obtained by an unweighted least-squares fitting of the PSF or LSF to the data. This is slightly suboptimal in terms of precision, but eliminates the risk of a magnitude-dependent bias.

The response functions that have to be calibrated are the two-dimensional PSF \( P_k(p, q) \) for the ASM, AF and BBP, and the one-dimensional LSF \( L_k(p) \) for the AF, BBP and MBP. Here, \( p \) and \( q \) are continuous relative pixel coordinates along and across scan, respectively. The functional representations must respect the sampling property (\( \sum_n L_k(p + n) = 1 \) for any \( p \), etc.) as well as the band-limitedness in the frequency domain (from diffraction).

In the presence of chromaticity, the centroid position depends on the spectral composition of the source. In the model adopted for algorithm 1-07-C, the image in each of the broad bands is displaced \( \Gamma_m \), producing a displacement of the total image of \( \sum_m \Gamma_m \phi_m \), where the chromaticity coefficients are constrained by \( \sum_m \Gamma_m = 0 \). A positive \( \Gamma_m \) means that the light in band \( m \) has an early (TBC) transit time, as compared to the average of the other bands. The chromaticity will introduce a certain displacement between the images in each of the bands, but the displacement between one of these images and the centroid position for the observation will depend on the spectral composition of the star. It amounts to \( \gamma_{ki} = \Gamma_{km} - \sum_{m'=1}^{M} \phi_{im'} \Gamma_{km'} \). It is not foreseen to apply such a correction, except for the AF CCDs.

\(^{12}\) That is, the value \( p_0 \) minimising \( \int [L(p_0 + p) - L(p_0 - p)]^2 dp \).
A.4 Response function (LSF/PSF) calibration

In a wide band CCD, also the shape of the response function depends on the spectral composition of the light, e.g. as given by the normalised BBP fluxes $\phi_{im}$. This can be taken into account by decomposing the LSF into $M$ quasi-monochromatic response functions, as in:

$$L_k(p) = \sum_{m=1}^{M} \phi_{im}L_{km}(p + \gamma_{ki}).$$  \hspace{1cm} (13)

The parameterisation of $L_{km}(p)$ follows rather closely the proposal in GAIA–LL–046:

$$L_{km}(u) = \frac{a_0}{1 + (u/a_2)^2} + \sum_{i=1}^{n} b_iB_i(u),$$  \hspace{1cm} (14)

where

1. a fraction, $h$, of the signal is modelled by a Cauchy distribution
2. $a_0 = 2h^2Dp_y/\lambda_m$ is the wing profile height ($D$ is the aperture along scan, and $p_y$ is the pixel size [radians] along scan)
3. $a_2 = \lambda_m/(2h\pi Dp_y)$ is the wing half width
4. $b_i, i = 1 \ldots n$ are the unknown coefficients, $\sum b_i = 1 - h$
5. $B_i(u), i = 1 \ldots n$ are the ‘bi-quartic B-splines’ defined in GAIA–LL–046. The function $B_i(u)$ is non-zero between $u_{i-3}$ and $u_{i+3}$.
6. the B-splines are defined on the grid $u_i = u_0 + 0.5i$, (with $u$ measured in pixels)
7. $u_0$ is initially $-(n + 1)/4$, but is fine tuned after the last iteration when the final solution for the $b_i$’s has been reached, so that the centroid is at position $u = 0$ (geometrical normalisation).

For each $L_{km}$ we therefore need to determine the $n$ coefficients, $b_i$, and the coordinate origin, $u_0$. The fraction $h$ is chosen at run time and is common to all the $L_{km}$. The number of B-splines, $n$, should be roughly twice the number of pixels over which the core of the LSF may be fitted.

The observation equation for $b_{kmx}$ resulting from sample $s_{lj}$ within observation $l$ is

$$\sum_{m=1}^{M} f_{lm}\phi_{im}(l)m \left[ \frac{a_{0m}}{1 + ((p_j + \gamma_{ki})/a_{2m})^2} + \sum_{x=1}^{n} b_{kmx}B_x(p_j + \gamma_{ki}) \right] \sim s_{lj} - b_l$$  \hspace{1cm} (15)

where

$$f_l = \sum_{m=0}^{M} (A_{km} + a_{km})g_{i(l)m}$$  \hspace{1cm} (16)

is the predicted flux in the observation and $p_j$ the calculated phase (in pixels) of sample $s_{lj}$ relative to the centroid.

The PSF $P(p,q)$ depends strongly on the across-scan motion of the image, which varies between 0 and twice the diffraction FWHM. Either the PSF needs to be calibrated as
function of across-scan velocity (which is of course accurately known from the attitude),
or it is calibrated only for small velocities and numerically modified as function of velocity.
A similar consideration concerns the along-scan smearing of the LSF due to TDI mismatch,
although that effect is on sub-pixel level.

The double star analysis requires the full two-dimensional PSF for accurate fitting of
overlapping components. The PSF is determined by similar principles as the LSF.

In the above presentation the observed flux is modelled in terms of $(M+1)$ standard
fluxes, the LSF is decomposed into $M$ quasi-monochromatic response functions, and the
chromaticity is also expressed as $M$ terms. All this fits the needs for the AF CCDs.
For the BBP CCDs we only need $M$ standard fluxes for the observed flux, one response
function for the LSF and no chromaticity terms. It is less clear what should be done
for the MBP CCDs, but probably three MBP bands for modelling the observed flux, one
response function for the LSF, and again no chromaticity terms.

### A.5 Photometric calibration

The basic model for the individual samples $s_{ln}$ in observation $l$ is

$$b_l + f_l \sum_{m=1}^{M} \phi_{i(l)m} L_{k(l)m} (p_n + \gamma_{k(l)i(l)}) \sim s_{ln},$$

where $p_n$ the calculated phase (in pixels) of sample $s_{ln}$ relative to the centroid and
$L_{km}(p)$ is the LSF given by (??). A chromaticity correction, $\gamma_{ki} = \Gamma_{km} - \sum_{m'=1}^{M} \phi_{im'} \Gamma_{km'}$, to $p_n$
is also included for the AF CCDs. Given $p_n$, the normalised standard fluxes $\phi_{im}$, and the
calibration parameters for the LSF, it is possible to estimate the background and source
fluxes, $b_l$ and $f_l$, from the above equation. (Additional data may be used to determine $b_l$.)
This estimation is performed by the centroiding routine, and is not part of the photometric
calibration.

The resulting ‘observed’ fluxes $f_l$ for the (non-variable) source $i = i(l)$ need to be modelled
in terms of the standard fluxes $g_{im}$ uniquely related to the magnitudes $G_i$, $B1_i$, etc. A
simple linear model is adopted, viz.:

$$f_l = \sum_{m=0}^{M} \left( A_{k(l)m} + a_{\kappa(l)m} \right) g_{i(l)m}.$$

For the Astro calibrations we take $m = 1 \ldots M$ to represent the BBP bands as before. For
calibrating the BBP CCDs this is all we need, but for the AF and ASM CCDs we add the
band $m = 0$ representing the $G$ band.

For Spectro calibrations band $m = 0$ represents the band being calibrated, and bands
$m = 1 \ldots M$ a number (two ?) of additional MBP bands.

Again, $A_{km}$ are the large-scale photometric parameters and $a_{\kappa m}$ the small-scale photometric parameters.
A.5 Photometric calibration

As suggested in Sect. A.2, standard fluxes \( g_{im} \) (and hence the instrument system) are defined through mission averages of the relevant response functions. For a hypothetical source, observed on all large scale units, this means

\[
g_{im} \equiv \langle f_l \rangle_{l; i(l) = i \land m(l) = m}.
\]

In practise, this average is obtained through the normalisation of the calibration coefficients. It can be noted that it is rather trivial to restrict the normalisation to a certain period of the mission, say the first year, rather than the whole mission, if it is desirable to obtain a stable system before the mission is completed. In analogy with (6) we separate the large and small-scale calibration parameters by means of the constraint

\[
\langle a_{\kappa m} \rangle_{\kappa \in K_k} = 0 \quad \text{for all } k, m.
\]

The large-scale photometric calibration parameters obey the orthogonality constraints

\[
\langle A_{km'} \rangle_{k; m(k) = m} = \delta_{mm'}.
\]

It is then possible to determine standard fluxes and the calibration parameters through iteration of the observation equations obtained from (18) in analogy with (8)–(10):

\[
\sum_{m=0}^{M} (A_{k(l)m} + a_{\kappa(l)m}) g_{i(l)m} \sim f_l
\]

\[
\sum_{m=0}^{M} a_{\kappa(l)m} g_{i(l)m} \sim \hat{f}_l - \sum_{m=0}^{M} A_{k(l)m} g_{i(l)m}
\]

\[
\sum_{m=0}^{M} A_{k(l)m} g_{i(l)m} \sim \hat{f}_l - \sum_{m=0}^{M} a_{\kappa(l)m} g_{i(l)m}
\]