The astrometric data-reduction software for exoplanet detection with PRIMA


Event: SPIE Astronomical Telescopes + Instrumentation, 2008, Marseille, France
The Astrometric Data-Reduction Software for Exoplanet Detection with PRIMA

Nicholas M. Elias II\textsuperscript{a,b}, Rainer Köhler\textsuperscript{a,b}, Ingo Stilz\textsuperscript{a,b}, Sabine Reffert\textsuperscript{a}, Ronny Geisler\textsuperscript{a,b}, Andreas Quirrenbach\textsuperscript{a}, Jeroen de Jong\textsuperscript{c}, Françoise Delplancke\textsuperscript{c}, Robert N. Tubbs\textsuperscript{b}, Ralf Launhardt\textsuperscript{b}, Thomas Henning\textsuperscript{b}, Denis Mégevand\textsuperscript{d}, and Didier Queloz\textsuperscript{d}

\textsuperscript{a}Zentrum für Astronomie der Universität Heidelberg, Landessternwarte; Königstuhl 12; 69117 Heidelberg; Germany;
\textsuperscript{b}Max-Planck-Institut für Astronomie; Königstuhl 17; 69117 Heidelberg; Germany
\textsuperscript{c}European Southern Observatory; Karl-Schwarzschild-Strasse 2; 85748 Garching bei München; Germany
\textsuperscript{d}Observatoire Astronomique de l’Université de Genève; 51, chemin des Maillettes; CH-1290 Sauverny; Switzerland

ABSTRACT

PRIMA/PACMAN is scheduled for commissioning on Paranal in late 2008 as part of the VLTI. In this paper, we discuss the important aspects of its astrometric data-reduction software. For example, the top-level requirements, interfaces to existing ESO software, data types, data levels, and data flow among the recipes dictate the overall design of any software package. In addition, the complexity of the PACMAN instrument, the long-term nature of astrometric observations, and the need to improve algorithms as the understanding of the hardware improves, impose additional requirements on the astrometric data-reduction software.

Keywords: exoplanet detection, parallax, proper motion, PRIMA, PACMAN, VLTI, ESO, narrow-angle astrometry, optical interferometry, data-reduction software

1. INTRODUCTION

The PRIMA (Phase-Referenced Imaging and Micro-arcsecond Astrometry) facility at ESO/VLTI will be commissioned on Paranal during the end of 2008. It will have two operational modes, phase-referenced imaging and dual-star narrow-angle astrometry.\textsuperscript{1,2} The Exoplanet Search with PRIMA (ESPRI) consortium, which is constructing parts of the instrument in collaboration with ESO, will use guaranteed time observations (GTO) to perform an astrometric exoplanet survey.

The astrometric mode is called PACMAN. The stars must be separated by less than \( \approx 30 \) arcsec, the size of the isoplanatic patch at K band. The final results are the fully calibrated differential delays, narrow-angle baselines, and projected angular separations on the sky. These quantities can be used to fit for parallaxes, proper motions, exoplanet orbits, etc. The projected angular separations must be accurate at the \( \approx 10-30 \) \( \mu \)as level to detect Saturn-sized exoplanets at distances of \( \approx \) tens of parsecs.

PRIMA consists of the following subsystems – two star separators (STTs), a differential laser metrology system (PRIMET), two fringe sensor units (FSUs), and four differential delay lines (DDLs) \textit{in vacuo} – that are used in conjunction with the existing auxiliary telescopes (ATs) and main delay lines (MDLs).\textsuperscript{1,2} One STS is required for each AT. In normal mode, FSU A is used for the program (normally bright) star, while FSU B is used for the reference (normally faint) star. In swapped mode, the roles are reversed. Calibration of systematic errors in the presence of random errors is key for the success of PRIMA and PACMAN.\textsuperscript{3}

PRIMET is comprised of two separate laser interferometers, PRIMETA and PRIMETB, operating at \( \approx 1.3 \) \( \mu \)m. They correspond to FSU A and FSU B. Each interferometer measures the change in optical path length (OPL) from telescope to beam combiner for each star light beam. PRIMET = PRIMETA - PRIMETB measures...
the differential delay between two stars. In single-star mode, PRIMETB records the delay for a single star, which is used to estimate the initial (wide-angle) baseline. It also acts as a constant-term metrology system, monitoring motions within VLTI.

The FSUs monitor stellar fringes in five narrow ($< 0.016 \, \mu m^{-1}$) channels spanning the Johnson K filter ($\approx 0.4-0.5 \, \mu m^{-1}$). Most interferometers track fringes by “dithering” across the central fringe. Each dither cycle is typically divided into four bins (labeled $ABCD$) that are separated by $90^\circ$. Visibilities and phases may be calculated directly from these bins. PRIMA is unique among interferometers because it employs a polarization modulation scheme that yields all $ABCD$ bins simultaneously with no dithering.\textsuperscript{4,5} Figure 1 contains a schematic diagram.

![Figure 1. The schematic diagram of an FSU, showing the achromatic quarter-wave plate, glass compensator, beam combiner (BC), and polarizing beam splitters (PBSs). This figure comes from Murakawa and Mathar.\textsuperscript{4}](https://ebooks.spiedigitallibrary.org/conference-proceedings-of-spie)

The astrometric data-reduction software (ADRS) is responsible for processing the raw data from these various subsystems, ultimately producing science-grade data products. The Landessternwarte (LSW) team, within the ESPRI consortium, is responsible for writing and testing the ADRS. The LSW team will also work with ESO on integrating the ADRS with the rest of the instrument. In addition, the ESPRI consortium will create the initial versions of semi-permanent calibration files.

## 2. TOP-LEVEL SOFTWARE REQUIREMENTS

Astrometry, by definition, is a medium- or long- term endeavor because of the motion of the sun with respect to other stars, the perturbed motion of the earth around the sun, the orbits of exoplanets, etc. Therefore, medium- or long- term systematic errors must either be minimized or removed by the instrument design, observational techniques, and the ADRS algorithms, otherwise false motions on the celestial sphere will mimic real astrometric behavior.

The ADRS must be simple enough to achieve the accuracies specified in the error budget.\textsuperscript{3} It must also be expandable, because we expect to improve the algorithms over time as we learn more about the performance of a new type of VLT instrument. This statement is especially relevant during the commissioning and “shared-risk” phases. Both simplicity and expandability require a good initial design so that future ADRS versions do not become morasses of spaghetti code.
Continuing checks of health and performance are necessary for all astronomical instruments. Some metrics (criteria) are very generic. For example, laboratory darks can show us if a pixel suddenly becomes “hot” compared to the others. Since PACMAN consists of two interferometers used in tandem, interferometer metrics should be applicable as well. Recording the total number of photons (or, the sum of $ABCD$ phase bins) tells us about the photometric stability of the atmosphere and/or instrument. The squared visibility and the phase variance are indicators of seeing and/or fringe-tracking quality. A constant differential phase over an entire scan (when both stars are bright) proves that the relative fringe tracking between FSUs is stable.

On-line processing, which is performed on Paranal (responding to the instrument in real time) and later by users at their home institutions, consists primarily of simple averaging. Almost all calibration is postponed until off-line processing, because we expect that most of the continuous algorithm development mentioned above will be related to removing systematic errors. This strategy has an additional benefit, namely that under most circumstances on-line processing needs to run only once. Recipe descriptions may be found in Sections 3 and 6.

Command-line parameters for the on-line processing must have meaningful defaults for the real-time system on Paranal, since operators will be too busy to manually change them before each observation. During the day, either the user (visitor mode) or day astronomer (service mode) will assess the data quality. Then, the user or day astronomer will run the off-line processing. If previous observations for a given star pair are available, they can be included in the analysis. Note that the day astronomer is not required to create science-grade data; that job is performed by the users at their home institutions.

In addition to nightly calibration files, the ESPRI consortium must create semi-permanent calibration files that are updated $\approx$ every six months. Some may originate in ASCII format, but they must be converted to FITS files before being used by the ADRS. Once the software and procedures have been developed for maintaining the calibration files, ESO takes on this responsibility.

The ADRS must be written in C, including embedded DOXYGEN documentation commands. Each algorithm corresponds to its own function, which simplifies incorporating additional improvements. Functions called multiple times are located in separate libraries. The ADRS must run under standard operating environments created and maintained by ESO. All non-scratch files must employ FITS format. If the ADRS requires non-standard tools, they must be approved by ESO.

3. INTERFACING WITH ESO SOFTWARE

VLT users create observation blocks (OBs) to run the instrument. Before acceptance, all OBs are verified by ESO. The OBs call templates that execute specific instrument tasks, such as acquiring a star, measuring a sky background, etc. Depending on the type of template, it may trigger an on-line recipe after completion that produces data products from raw data.

A recipe is a collection of related C functions in a single file that perform specific tasks. It provides a single external interface and encapsulates internal data. Both of these features represent elementary object-oriented programming techniques. A description of the ADRS on-line and off-line recipes may be found in Section 6.

As mentioned above, templates may trigger on-line recipes on Paranal. The recipes may also be run by users under operating systems created and maintained by ESO: gasgano and esorex. Off-line processing can only be run under these operating systems.

Gasgano executes recipes, but it does much more. Via a graphical user interface (GUI) it automatically associates related files to each other and feeds them to the desired recipe. For example, a file containing both raw fringe and metrology data can be associated with sky background and star flat files, and then all of them can be sent to the on-line processing recipe that computes averages.

Esorex does not have as many features as gasgano. For example, esorex has no GUI. Also, its primary input is an ASCII set of frames (SOF) file, which contains the path, name, and data organization (DO) category of each input file. The SOF files may be created manually, but it is more convenient to create them automatically.

The PACMAN ADRS has an off-line recipe, called the data analysis facility (DAF), which is unique among VLT instruments. It is used to fit differential delay and delay data to environmental data from the ESO engineering database. In turn, these fits are used to remove long-term trends. The recipe is used in conjunction...
with a program called OCA (organization, classification, and association) to create SOFs. OCA reads a set of rules to associate the environmental files and the science files. This scheme minimizes the modifications to software when a new trend is found. A diagram of this process may be found in Figure 2.

![Diagram of the DAF recipe and its interface to OCA]

Figure 2. The DAF recipe, including its interface to OCA.

4. DATA TYPES

PACMAN employs many different types of data. In Figure 3, we present a data tree showing how all of the data flow up to the final result, the angular separation projected on the sky.

The most basic PACMAN calibration information are the laboratory dark and sky background. At present, we do not subtract lab dark from sky background, which is acceptable if the latter have the same detector integration time (DIT) as the files they calibrate. The laboratory darks are used in conjunction with interferometric Fourier transform spectroscopy (FTS) measurements of a 600 °C laboratory source (MARCEL) to yield the complex FSU bandpass versus wavenumber. Similarly, sky background is combined with other FTS measurements to provide VLTI bandpass and stellar SED versus wavenumber. All of these results are used to calculate the effective wavenumber \( \kappa_{\text{eff}} \) for each pixel and the effective \( ABCD \) phase deviations from their nominal values \( \Delta \phi_{\text{eff}} \). The FSU bandpasses must be obtained before each night’s run, while the VLTI bandpasses need to be formed \( \approx \) every six months.

The sky background and star flat (relative pixel gains) are fed into the recipe that converts fringe data into interferometric observables, such as the complex cross visibility \( V_x(\kappa) = V_A(\kappa)V_B^*(\kappa) \), which is the product of the complex visibilities measured by both interferometers. If the phase differences of the \( ABCD \) bins are not exactly 90°, “stepping” may be employed to remove these effects from the cross visibilities.\(^5\) Note that stepping is not exactly the same as dithering; the former employs discrete delay steps while the latter is continuous. The cross visibility phases are used in conjunction with the effective wavenumbers and effective \( ABCD \) phase deviations to calculate the offset corrections for the differential delays recorded by PRIMET.
We expect the differential delays to be affected by environmental conditions. These data, obtained from the ESO engineering database, are Fourier filtered and interpolated onto the same average time grid as the fringe data in order to eliminate the need for additional interpolation later in the data reduction process. The filtering is required because it is unlikely that high-frequency components will be correlated with fringe data. The processed environmental data are then combined with environmental fit parameters that are stored in the semi-permanent calibration files called the correction collection (CoCo). These parameters are determined by fitting the differential delay data of calibration sources versus environmental data. They are updated on ≈ a six month schedule.

In addition to the desired astrometric motions, others such as diurnal/annual aberration, light deflection by solar system bodies, light time delay, earth orientation, etc., appear in the differential delays. The latter dwarf the former, so their effects must be removed. Stellar positions are required for these calibrations, and they are stored in catalogs in the CoCo. Unfortunately, no star catalog contains positions accurate enough for our purposes. When PACMAN is used to observe the same star pair multiple times, it will be possible to calculate corrections to the position differences between the stars. This information, in conjunction with the initial catalog positions, is sufficient for calibration.

The PRIMET metrology system has an unknown offset. Also, it is possible that the differential constant
term (the difference between the constant terms of interferometers A and B) exhibits a slow linear drift over \( \approx 30 \) minutes, the time required for a complete observation of a star pair. To remove these effects, we obtain four \( \approx 2.5 \) minute stellar scans per observation. In the two middle scans, the STSs “swap” the stars to the other interferometer. With such an ensemble of data, it is possible to simultaneously fit for the astrometric motions, PRIMET zero point, and differential constant-term drift.

The effective wavenumbers and the processed environmental data are used to calculate the “dispersion” correction for the PRIMETB delays. This term is somewhat of a misnomer. Granted, dispersion of star light does manifest itself in the delays, but it is too small for effective corrections, which means that no corrections for slow delay drifts due to atmospheric turbulence can be calculated. The VLTI main delay line, which is supposed to compensate for the geometric delay above the atmosphere, is filled with air. Fortunately, PRIMETB is stabilized \textit{in vacuo}, which means that we can use the Mathar model\(^\text{6}\) of the refractive indices to eliminate the effects of this unwanted air.

Once the “dispersion” correction has been applied to the delays, they are fed into the IPhASE library, which calculates the initial (wide-angle) baseline. The required accuracy for the star positions is on the order of 100 mas, which can be obtained with the FK6 catalog\(^*\). The narrow-angle baseline tracks how differential delays change as a star pair moves across the sky.\(^3\) For identical telescopes and feed systems, the wide- and narrow-angle baselines are the same. The exact nature of the corrections for real telescopes and feed systems will be determined in the near future.

### 5. DATA LEVELS

PACMAN generates science and calibration data. Both forms exist as either raw data or processed data products. All FITS file formats are located in the ADRS design document.\(^7\)

The science data are divided into levels. One type of calibration data, environmental data from the ESO database, also has its own levels. Each level represents the amount of ADRS processing. There are six of them:

- **Level 0**: Raw fringe and delay science data, obtained at ms rates.
- **Level 0e**: Raw environmental data from the ESO engineering database.
- **Level 1**: “One-second” (exact timing value TBD) averages and errors of level 0 data.
- **Level 1e**: Filtered level 0e data interpolated onto the level 1 time grid.
- **Level 2**: Instrument corrected (environment, offset/dispersion) “one-second” main-delay averages, differential-delay averages, PRIMET FSUB averages, and their errors.
- **Level 3**: Fully corrected differential delays, projected sky angles, narrow-angle baselines (lengths and orientations projected on the sky), and their errors.

Apart from the separation of level 1/1e and level 2 data, this sequence represents the canonical view of data reduction for most astronomical interferometry: averaging, corrections, and final data product. Note that the ADRS performs no orbit fitting. Creating or obtaining such software is the user’s responsibility.

Each level 0 file is \( \approx 500 \) Mb in size and contains all of the raw science FITS tables. The level 0 data types include FSU, MDL (main delay line), PRIMET, PRIMETB, etc. The level 0e data are collected over 24 hours into a single FITS file at the end of each night. It contains temperatures, pressures, relative humidities, tip-tilt, pointing, etc. The format may be found in the data interface control document (DICD).\(^8\)

Each level 1 file corresponds to a single level 0 file and has FITS tables with considerably fewer data. The level 1 file name prefixes are the same as the level 0 file name prefixes except that the former have “LEVEL1” appended to them. Averaging means different things for different types of data in the file. Metrology data are averaged by performing linear fits over the raw data corresponding to each one-second time stamp. Fringe data are converted to interferometric quantities, such as squared visibilities and cross visibilities, and averaged simply.

\(^*\)http://www.ari.uni-heidelberg.de/datenbanken/fk6/index.php.en
Similarly, each level 1e file corresponds to a single level 1 file. The level 1e file name prefixes are the same as the level 1 file name prefixes except that the former have “LEVEL1E” appended to them. The level 1 and 1e FITS files contain header quality control (QC) keywords, such as minimum/maximum values, minimum/maximum standard deviations, scan averages and standard deviations, etc. The averaging process involves Fourier filtering and interpolation onto the level 1 time average grid.

Level 1, level 1e, nightly calibration, and CoCo files are used to produce a level 2 file. The level 2 file name prefixes are the same as the level 1 file name prefixes except that the former have “LEVEL2” appended to them. The level 2 data are used to create the level 3 data. Three level 2 files from two different swap states (unswapped, swapped [twice as long], unswapped) and CoCo files are used to create a single level 3 file, corresponding to a single stellar observation (not necessarily the same as an OB). The level 3 file name prefix is the same as the first level 0 file name prefix for each observation plus “LEVEL3”. The level 2 and 3 FITS files contain header QC keywords, such as minimum/maximum values and minimum/maximum standard deviations, scan averages and standard deviations, etc.

Nightly calibration files include lab darks, sky backgrounds, star flats, and FTS data for FSU throughputs. These data types are similar to the data types for other VLT instruments. CoCo calibration files include star catalogs, FTS data for VLTI throughputs and star SED measurements, star SED templates, and environmental fit parameters. These data types are unique to PACMAN, since medium- to long- trend calibration is required for astrometric programs.

6. RECIPES AND DATA FLOW

There are two types of ADRS recipes, on-line and off-line. The on-line recipes can respond to templates in real time on Paranal or be run by users at their home institutions. In service mode, the off-line recipes are run by the day astronomer after a night of observations. In visitor mode, the user reduces the data himself/herself. Depending on the amount of effort, science-grade results are possible on Paranal, if desired.

The on-line recipes are:

lab dark: This recipe processes a raw laboratory dark calibration file, containing multiple exposures with different integration times, and creates an averaged file with biases and dark currents for each pixel.

lab flat: This recipe processes a raw laboratory flat calibration file and creates a file of relative pixel gains. It is not used as part of the standard ADRS reduction, it is used for laboratory health checks.

fsu response: This recipe processes FTS data of a laboratory source (MARCEL) and creates a file of complex FSU bandpasses for each pixel.

sky background: This recipe processes a raw sky background calibration frame and creates a frame of averaged sky backgrounds.

vlti response: This recipe processes FTS data of a star and creates a file of complex VLTI bandpasses for each pixel.

star spectrum: This recipe processes FTS data of a star and creates a stellar SED. Standard SED files in the CoCo can be used in place of this recipe.

star flat: This recipe processes a raw star flat calibration file and creates a file of relative pixel gains. Unlike the lab flat, it is part of the standard ADRS data reduction.

sciave: This recipe calculates “one-second” processed data from raw laser metrology data (linear fitting) and fringe data (averaging).

The complex bandpasses, determined via FTS, contain the real and imaginary parts of the throughputs versus wavenumber. They are used to estimate \( \kappa_{\text{eff}} \) and \( \Delta \phi_{\text{eff}} \) for each pixel. The association map for the on-line processing may be found in Figure 4.

The off-line recipes are:
**environment:** This recipe processes raw environmental data from the ESO engineering database by low-pass filtering and interpolating them onto the average time grid created by sciave.
This recipe calculates the effective wavenumber and effective phase offsets from the FSU bandpass, VLTI bandpass, and star SED; calculates offset corrections for the differential delays from the cross-visibility phases; calculates “dispersion” corrections using environmental data and a refractive index model; and calculates environmental corrections for both the differential delays and delays.

**baseline:** This recipe calculates the initial (wide-angle) baseline estimate from the corrected delays using the IPHASE library.

**scired2:** This recipe calculates the narrow-angle baseline correction, removes the arbitrary PRIMETB zero point and slow drift of the differential constant term from the differential delays, removes unwanted astrometric effects from the differential delays, and calculates the angular separations on the sky.

**daf:** This recipe is the data analysis facility (DAF). It is not part of the standard ADRS data reduction. The ESPRI consortium and ESO use it to create environmental fit parameter files and save them to the CoCo.

The association map for the off-line processing may be found in Figure 4. The DAF is not included because it is not meant to be run by users other than the ESPRI consortium and ESO.

## 7. CONCLUSION

ESO and the ESPRI consortium will start PACMAN commissioning on Paranal toward the end of 2008. The process will test not only the hardware but the ADRS software interfaces and algorithms. The complexity of PACMAN observations means that the overall ADRS design must take into account the long-term nature of astrometric observations and the need to improve algorithms as the understanding of the hardware improves.

In this paper, we presented a brief description of the PACMAN components and their purposes. We also described the ADRS top-level requirements and how the ADRS interfaces with existing ESO software. Last, we discussed the ADRS data types, data levels, and data flow among recipes.

## ACKNOWLEDGMENTS

RG is an IMPRS fellow of Astronomy and Cosmic Physics at the University of Heidelberg.

## REFERENCES


