



Vera C. Rubin Observatory
Data Management

Astrometric Calibration in the LSST Pipeline

Clare Saunders

DMTN-266

Latest Revision: 2024-10-02



Abstract

This technical note will describe our procedure for fitting an astrometric solution in the LSST Data Release Production.

Change Record

Version	Date	Description	Owner name
1	2024-10-02	Initial release.	Clare Saunders

Document source location: <https://github.com/lsst-dm/dmtn-266>

Contents

1 Introduction	1
2 Astrometry in the science pipeline	1
2.1 Preliminary Astrometric Solutions	1
2.2 The Final Astrometric Solution	2
2.3 Apparent motion of objects	4
2.3.1 Differential Chromatic Refraction	6
3 Fitting Modes	11
3.1 Multiband fitting	11
3.2 "Global" fitting	11
3.3 Building and Using A Camera Model	12
4 Outstanding issues	14
A References	17
B Acronyms	18

Astrometric Calibration in the LSST Pipeline

1 Introduction

The goal of the astrometric fitting in the LSST science pipeline is to obtain precise alignment of individual exposures relative to each other along with accurate alignment with an external reference catalog. This alignment is necessary to produce accurate coadded images downstream in the pipeline. It also is necessary for obtaining precise measurements of all objects and for measuring proper motion and parallax. In particular, the astrometric calibration is necessary to attain the design requirements enumerated in the LSST System Science Requirements Document (LPM-17) and related documents.

2 Astrometry in the science pipeline

2.1 Preliminary Astrometric Solutions

An initial estimate of the World Coordinate System (WCS) mapping is made using the camera model. This camera model is currently assumed to be static (see Sections 3.3 and 4 for discussion of improvements to this approach). In single-frame processing, this WCS is further refined by matching point sources to an external catalog and fitting a mapping per detector for each exposure. This fit is done by `AstrometryTask` as part of `calibrateTask` or `calibrateImageTask`, (the latter will be the long-term place for this step). For this mapping, the map can be configured to be either a TAN-SIP WCS (Shupe et al. 2005) (implemented in `FitTanSipWcsTask`), or the input camera model followed by an affine transformation (`FitAffineWcsTask`). The latter is preferred from a perspective of speed and robustness, but will give poor results if the input camera model is not sufficiently accurate. `FitTanSipWcsTask` is used for most instruments in the DRP pipeline, whereas `FitAffineWcsTask` is used for the AP pipeline.

The reference catalog currently used is from Gaia DR3 (Gaia Collaboration et al. 2023), which covers the whole sky, with a limiting magnitude of about 21 in the Gaia G-band, which is analogous to the LSST r-band. The accuracy of this fit is limited by the number of sources in an exposure that can be matched to the reference catalog and by the uncertainty on the centroids of the sources in both the science and reference catalogs. However, after this initial fit, the WCS is accurate to within 1 arcsecond, which is sufficient for associating isolated sources

across visits.

For Prompt Processing, no further astrometric calibration is done. For the Data Release Production, this initial astrometric solution is further refined.

2.2 The Final Astrometric Solution

Following the single-frame processing, the final astrometric model is fit using an ensemble of overlapping visits. This fit is done in `gbdesAstrometricFitTask`, which runs the `wcsfit` fitter from the `gbdes` package originally developed for the Dark Energy Survey (Bernstein et al. 2017). The baseline running mode is to use all the visits that overlap with a given tract of the sky in one band. However, this fit can also be done using multiple bands and larger areas of the sky, or even globally. Running in these modes is discussed in Section 3.

In order to fit this model, isolated point sources are identified in all of the input images. Matches are then made among these sources, and between the sources from science images and the reference catalog. The principle that each of these matched sources corresponds to an object with some true position (as a function of time) is the base assumption that allows the astrometric model to be fit. Correspondingly, the fact that the reference catalog positions are on average in ICRS coordinates ties the astrometric model to absolute positions.

The source association is currently done either as part of the astrometric fitting task or externally in `IsolatedStarAssociationTask`. Using the external catalog of isolated stars is not possible when running the astrometric fit per-tract because of restrictions in the pipeline software. However, the long-term plan is to use the isolated star catalog in all cases. This will mean that the astrometric fit, the photometric fit, and other parts of the pipeline requiring isolated point sources will all use a consistent set of inputs.

The astrometric model itself is made up of a series of mappings, going from pixel-space to one or more intermediate frames, and then to sky coordinates. The most basic version of the astrometric model used in the Rubin Science Pipelines consists of two fitted maps, plus one that is deterministic from the telescope pointing: the fitted maps consist of one static map per detector that goes from the pixel coordinates for each detector to focal plane coordinates; then one map per visit that goes from the focal plane coordinates to coordinates on the plane tangent to the telescope boresight. These are followed by a map from the tangent plane to the sky, which is purely defined by the pointing. The intention with this is that the static per-

detector map will account for unchanging characteristics of the telescope and camera that affect the path of a photon up until collection in a pixel. In reality, some effects that are assumed to be static may vary on longer time scales, and this will be discussed in Section 4. The per-visit map then accounts for time-varying effects on the path of a photon—for example, refraction due to Earth’s atmosphere, atmospheric turbulence, and changes in the optics dependent on the position of the telescope. Both the per-detector and per-visit maps are modeled using a two-dimensional polynomial, generally on the order of 2 to 8. Complications to this picture will be discussed in Section 4.

In practice, the astrometric fit happens between pixel positions and the positions on a plane tangent to the approximate center point of the input visits. The model can then be expressed thus:

$$(x_{int}, y_{int}) = F_{detector}(x_{pix}, y_{pix}) \quad (1)$$

$$(x_{tp}, y_{tp}) = F_{visit}(x_{int}, y_{int}) \quad (2)$$

where F can be any two-dimensional function or a further composite of functions, (x_{int}, y_{int}) are coordinates in an arbitrary intermediate frame, and (x_{tp}, y_{tp}) are the coordinates in the plane tangent to the telescope pointing. The final step is a trivial transformation from the tangent plane to sky coordinates.

To fit the model, the χ^2 for all sources is minimized as a function of the mapping parameters, where the χ^2 function takes the following form:

$$\chi^2 = \sum_i ((x_{tp,m}, y_{tp,m})(t) - F_{visit} \circ F_{detector}(x_{pix,i}, y_{pix,i}))^T C_i^{-1} ((x_{tp,m}, y_{tp,m})(t) - F_{visit} \circ F_{detector}(x_{pix,i}, y_{pix,i})) \quad (3)$$

Where $(x_{pix,i}, y_{pix,i})$ is the pixel position for source i , and $(x_{tp,m}, y_{tp,m})(t)$ is the best fit position of the object m to which source i corresponds, and C_i is the covariance matrix for source i .

Including reference objects breaks various degeneracies in the model. The χ^2 term for reference objects with coordinate information only is

$$\chi^2 = \sum_i ((x_{tp,m}, y_{tp,m})(t) - F_{tp}^{-1}(RA_{sky,i}, Dec_{sky,i}))^T C_i^{-1} ((x_{tp,m}, y_{tp,m})(t) - F_{tp}^{-1}(RA_{sky,i}, Dec_{sky,i})) \quad (4)$$

or, for objects with proper motion and parallax information,

$$\chi^2 = \sum_i ((x_{tp,m}, y_{tp,m}, PM_{x_{tp,m}}, PM_{y_{tp,m}}, \pi_{tp,m})(t) - F_{tp}^{-1}(RA_{sky,i}, Dec_{sky,i}, PM_{RA,m}, PM_{Dec,m}, \pi_m))^T C_i^{-1} (x_{tp,m}, y_{tp,m})(t) - F_{tp}^{-1}(RA_{sky,i}, Dec_{sky,i}, PM_{RA,i}, PM_{Dec,i}, \pi_i)) \quad (5)$$

where F_{tp} is the (purely geometric) transformation between the tangent plane and the sky, $PM_{(RA,Dec),i}$ and π_i are the proper motion and parallax for the reference i in RA and Dec and $PM_{(x_{tp},y_{tp}),m}$ and $\pi_{tp,m}$ are the proper motion and parallax for the corresponding object in the tangent plane.

The model is fit by minimizing the χ^2 as a function of the mapping parameters. For each step of the minimization, the Hessian matrix of the χ^2 is calculated. With the Hessian fixed, the χ^2 is minimized by alternating between Newton iterations and recalculating the best-fit position (and apparent motion parameters; see Section 2.3 for further discussion) of the objects based on the mapping at that point. When these iterations cease to improve the χ^2 , outliers are clipped, and the Hessian matrix is recalculated. This is repeated until the improvement in the χ^2 is below a certain threshold.

Figure 1 shows the difference between the fit 6th-order detector model for the transformation between pixels and the field angle and a simple affine transformation for the HSC camera. Figure 2 shows the difference between the same 6th-order detector model and a second-order transformation, which highlights the existence of consistent behavior detector-to-detector.

2.3 Apparent motion of objects

One of the motivations for moving to the `gbdes` package was its ability to fit proper motion and parallax. Fitting these motions is currently optional in `gbdesAstrometricFitTask`, in order to allow for situations like fitting data that has been simulated without apparent motions, or fitting data with a very short time baseline where allowing apparent motion would add unwanted flexibility. In `gbdesAstrometricFitTask`, proper motion and parallax fitting are turned on by setting the configuration parameter `fitProperMotion=True`. The equations for the position of an object m is then

$$RA_m(t) = RA_{m,t_0} + PM_{RA,m}t + \pi P_{RA}(t, RA, Dec) \quad (6)$$

$$Dec_m(t) = Dec_{m,t_0} + PM_{Dec,m}t + \pi P_{Dec}(t, RA, Dec) \quad (7)$$

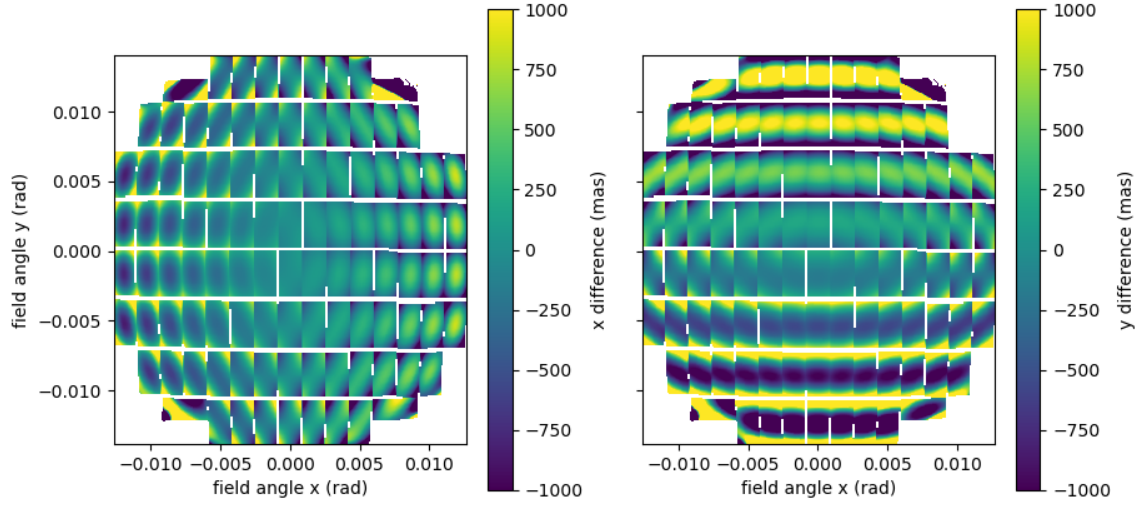


FIGURE 1: Comparison between the fit 6th-order detector model for the HSC camera and an affine (i.e. first order) model for the transformation between pixels and field angle.

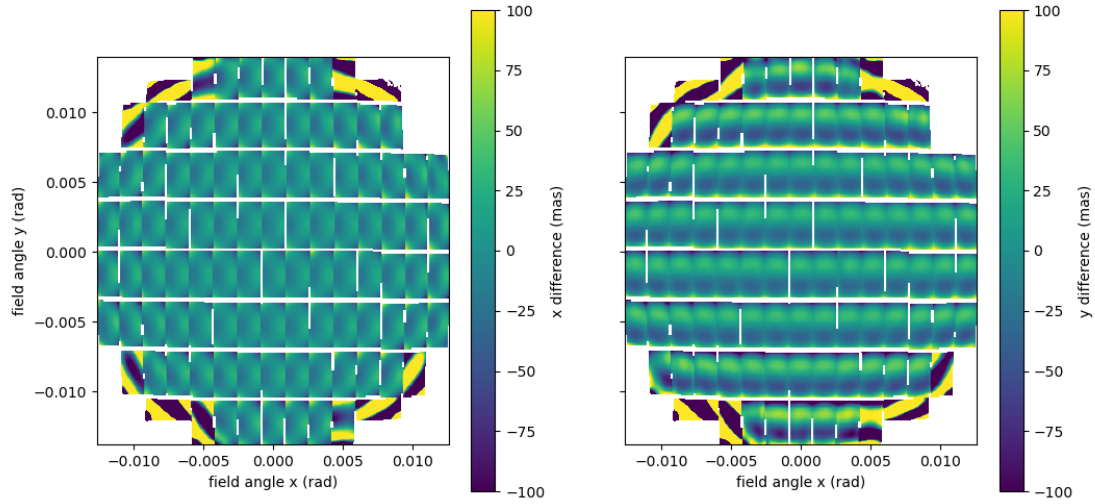


FIGURE 2: Comparison between the fit 6th-order detector model for the HSC camera and a second-order model for the transformation between pixels and field angle.

where $(RA, Dec)_m(t)$ is the observed position of object m at time t , $(RA, Dec)_{m,t_0}$ is the position at the reference epoch t_0 , $PM_{RA,Dec}$ is the proper motion of the object, π is the parallax of the object, and $P_{RA,Dec}$ is the parallax factor determined by the time t along with the (RA, Dec) of the object.

Figures 3 to 6 show the fits for proper motion and parallax compared to the GAIA DR3 values, along with the uncertainty in the fits as a function of magnitude, when fitting the COSMOS UltraDeep z-band data.

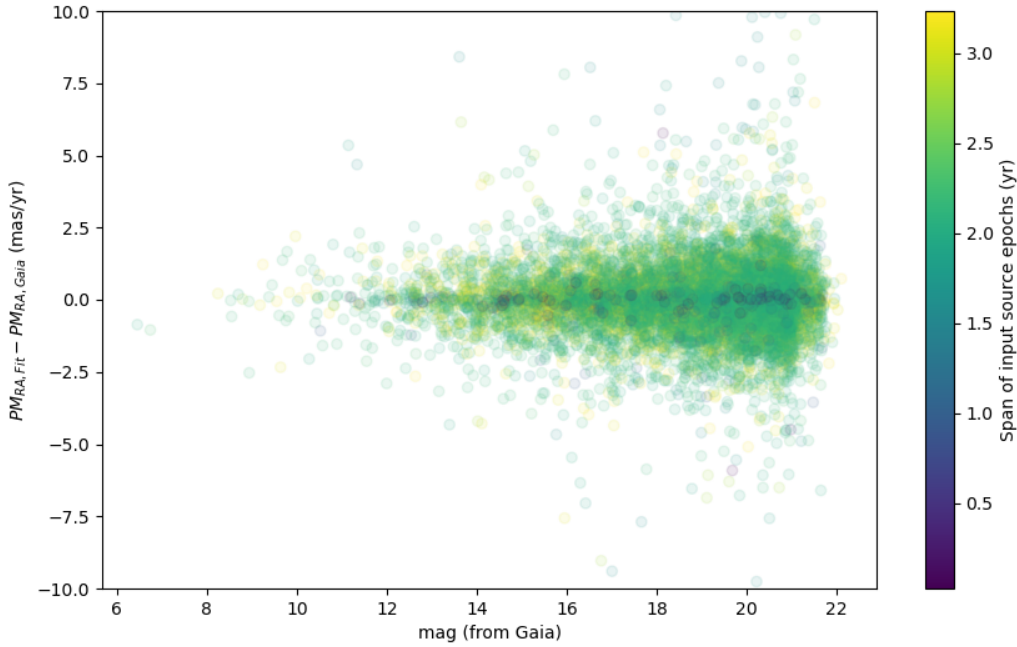


FIGURE 3: Comparison between the values fit for the proper motion (in RA) and the values from the GAIA DR3 catalog.

2.3.1 Differential Chromatic Refraction

Another optional correction can be made to the observed position of an object in order to account for Differential Chromatic Refraction (DCR). The Hyper Suprime-Cam data is not affected by this issue because of a color corrector, so this had not been a pressing focus for the development of the astrometric model. However, it is possible with the `gbdes` package, and has been tested on simulated data for ComCam. Turning on the color correction is done by

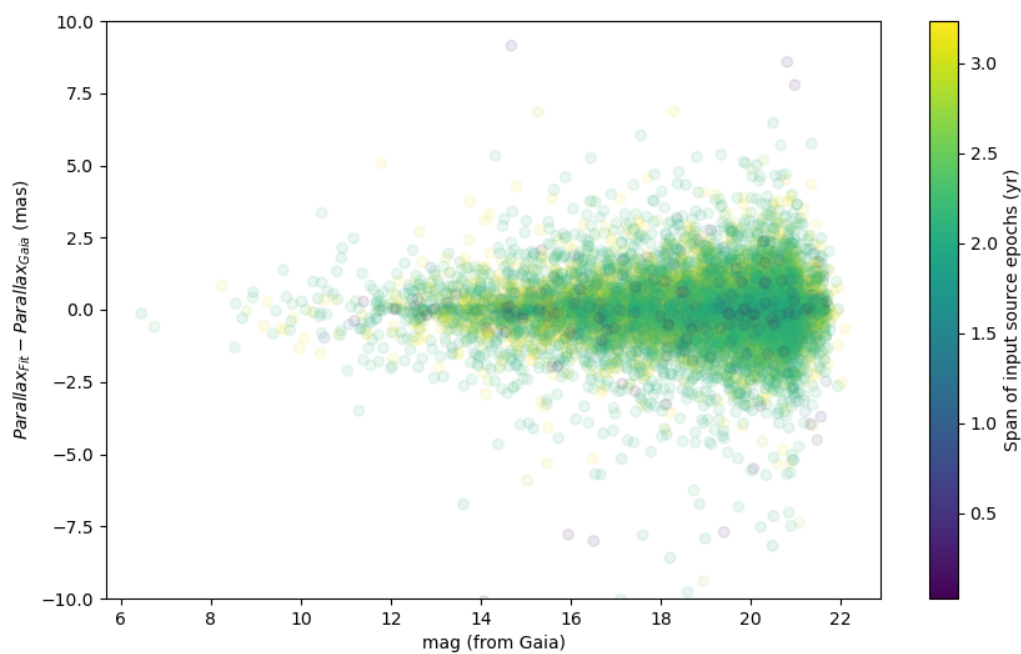


FIGURE 4: Comparison between the values fit for the parallax and the values from the GAIA DR3 catalog.

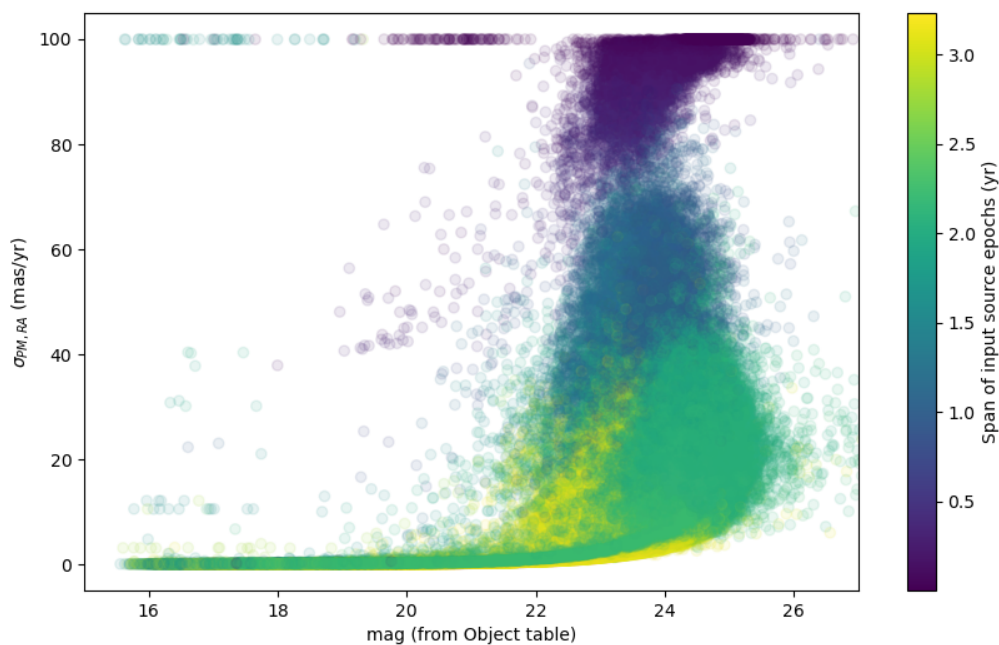


FIGURE 5: Uncertainty in the proper motion (in RA) fit as a function of the object magnitude. The color signifies the timespan between the first and last epoch of the input sources for each object.

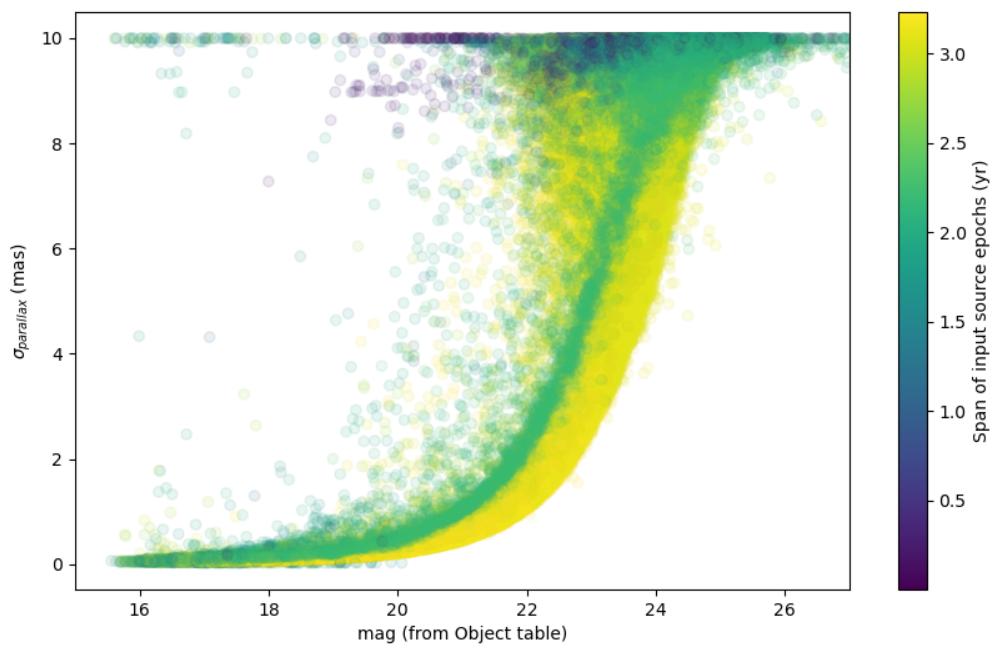


FIGURE 6: Uncertainty in the parallax fit as a function of the object magnitude. The color signifies the timespan between the first and last epoch of the input sources for each object.

setting the `gbdesAstrometricFitTask` configuration variable `useColor` to be `True`.

The correction currently implemented for DCR takes the form

$$\Delta \mathbf{x}_i = \mathbf{D}_{visit} c_m \quad (8)$$

where $\mathbf{D}_{visit}$ is the DCR coefficient determined by the pointing and atmospheric conditions of a given exposure and c is the color parameter for object m .

The main driver in variability of the DCR coefficient is the pointing, so $\mathbf{D}_{visit}$ can be modeled generally by the equation

$$\mathbf{D}_{visit} = \mathbf{D}_{band} \tan(z) \hat{\mathbf{p}} \quad (9)$$

where z is the zenith distance, $\hat{\mathbf{p}}$ is the unit sky vector towards zenith, and $\mathbf{D}_{band}$ is a constant per-band that is fit to the ensemble of observations. Figure 7 shows the linear relationship between the $\mathbf{D}_{visit}$ fit per visit and the pointing, using simulated ComCam data. However, given

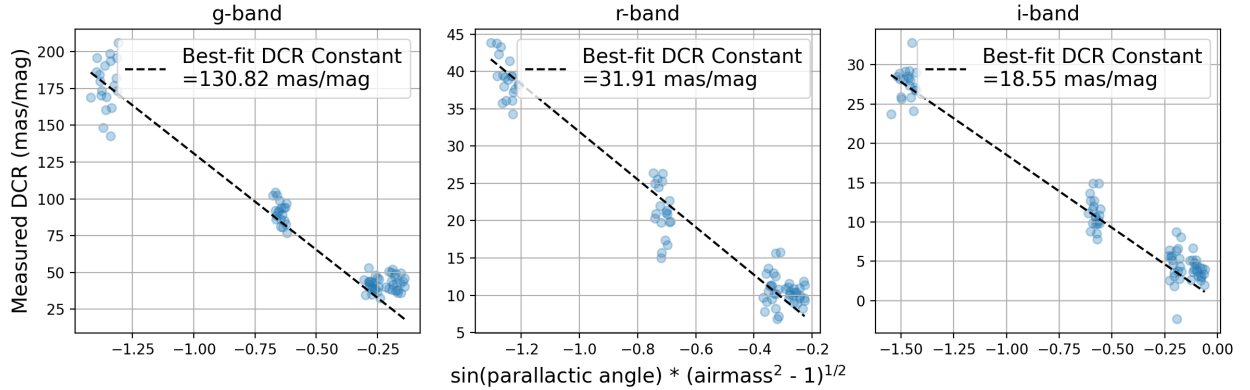


FIGURE 7: Fit values for $\mathbf{D}_{band}$ as a function of the telescope pointing.

the large number of sources per visit that are expected from LSST, we can likely get a better fit of the effects of DCR by fitting $\mathbf{D}_{visit}$ per exposure in order to allow for time-dependent atmospheric effects.

The default color used is $g - i$, but this is configurable in the task. The magnitudes used to calculate the color come from the `fgcmCa1` calibration of isolated stars, meaning that colors are not available for all objects due to selection differences.

3 Fitting Modes

The baseline mode for fitting the astrometric model, as described in Section 2.2, is to fit all of the visits in a given band that are overlapping with a given tract of the sky. However, the model fitting can objectively be improved by using larger amounts of data, and there are scenarios, which will be common for ComCam and in the early days of LSST, where there will not be enough data per tract in a given band. Because of this, a number of different running modes have been developed with an eye towards what will be needed during commissioning.

3.1 Multiband fitting

The astrometric model is expected to be different from band to band due to differences in transmission through both the atmosphere and the telescope optics and camera. However, the true positions of the sources are obviously the same regardless of the filter through which they are observed. Fitting multiple band simultaneously thus provides a stronger constraint on this aspect of the model and leads to a fit with greater precision. Running with multiple bands is done by the task `GbdesAstrometricMultibandFitTask`, which inherits from `GbdesAstrometricFitTask` and is algorithmically identical. Note that this is independent from the DCR fitting described in Section 2.3.1.

The multiband fit is currently significantly slower than in the single band task, most likely because the matrix that must be inverted in order to perform the fit scales with N_{band}^2 . However, the matrix in question is block diagonal, so improving the fitter by using this fact should improve the runtime problem, since the fit would then scale as N_{band} . Until this is implemented, the multiband running mode is only recommended in cases where there is a very small number of visits per band ($\lesssim 10$).

3.2 "Global" fitting

`GbdesGlobalAstrometricFitTask` is another class inheriting from `GbdesAstrometricFitTask` that adds the option to fit the astrometric model on sets of visits (almost) regardless of their location on the sky, rather than being limited to only visits overlapping with one tract. The advantage of this approach is that the static camera part of the astrometric model can be better constrained with more input visits. The planned use case for this task is not to actually fit the astrometric model for all available visits at the same time, but to allow for fitting the

static camera model on a selection of visits from potentially disjoint areas on the sky. By doing this, good-seeing data taken from a number of different fields on the sky (such as the deep-dithered star fields proposed for commissioning) can be used to fit the astrometric model. The static part of this model can then be used as a fixed detector model in subsequent fits of new input data from areas of the sky with potentially limited visits (see Section 3.3).

In practice, for the global task, the input visits are split into separate groups of overlapping visits. The extent of each of these sets of visits then becomes a separate “field”, which determines the tangent plane on which the astrometric fit is done. Because of this process, each group of visits must span less than a hemisphere, which is why `GbdesGlobalAstrometricFit` cannot actually be run globally in the scenario where the input visits actually cover the whole sky.

Because the task is not tract-based, it is possible to include the matched sources from `isolatedStarAssociationTask` as an input to the task. (See Section 2.2 for discussion of how this is not possible in the tract-based class.) Thus, it is not necessary to do source-matching in the task, and the resulting catalog of fit star positions is more consistent with that of other calibration tasks (like `fgcmCal`) that also use the `isolated_star_sources` as input.

Other than these two aspects of the task, the fit done in `GbdesGlobalAstrometricFitTask` is otherwise identical to that of `GbdesAstrometricFitTask`.

3.3 Building and Using A Camera Model

With a sufficient number of visits, the astrometric fit should be able to construct an extremely accurate model of the camera distortions from the static detector-based part of the `gbdes` model. During LSST, we may want to construct a camera model using good-seeing images on deep-dithered star fields. Using the “global” version of the astrometric fitter, these images could be fit together to provide a particularly strong constraint on the camera model. There are (at least) two potential uses for this model:

First, it can be used as input when performing the astrometric fit on tracts of the sky where there are a small number of visits. In this case, allowing both the per-detector and per-visit parts of the model to vary will result in overfitting. By using a per-detector model that has been trained on a large amount of data, and then holding that fixed, we can guard against this. In practice, this is done by running the fitter on some training set of visits with the configuration

option `saveCameraModel=True`, then fitting new data with the previous run as an input to the pipeline task and with the configuration option `useInputCameraModel=True`. Figure 8 shows the improvement in χ^2/dof for validations sources that were withheld from the astrometric fit when using an input camera model compared to letting the model be fit freely.

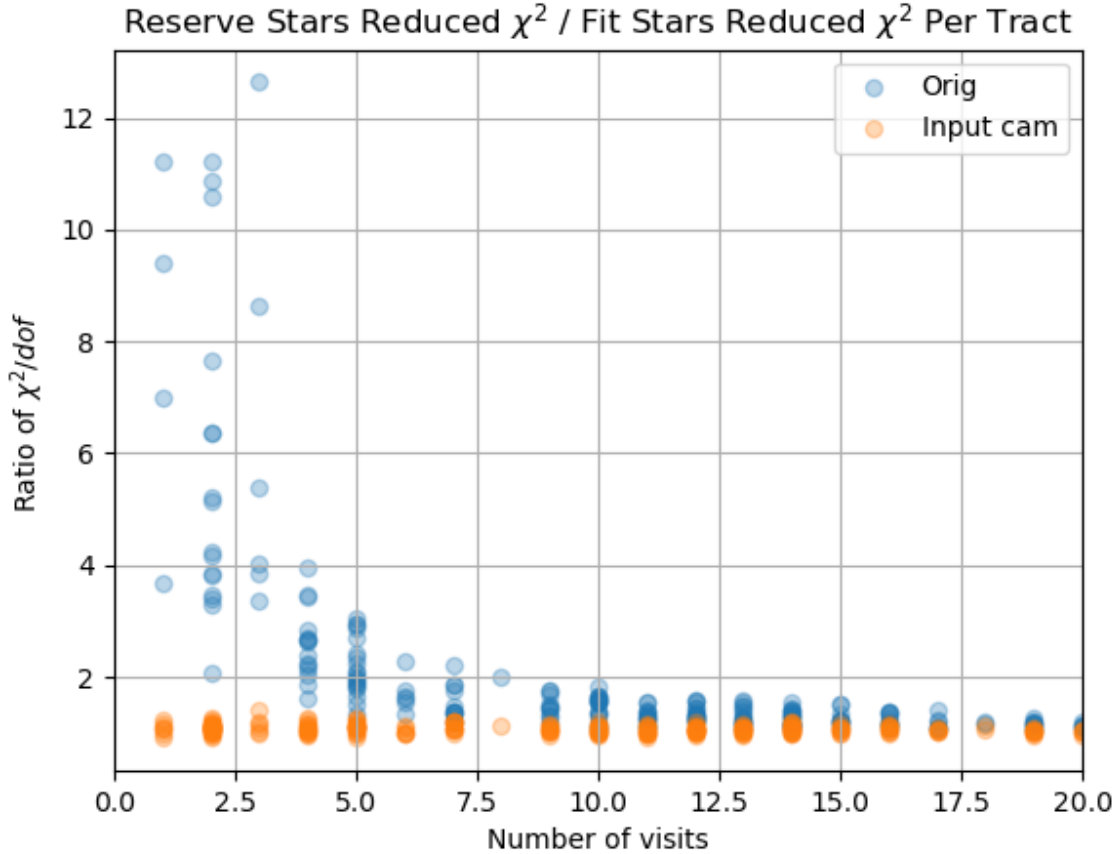


FIGURE 8: Ratios between the reduced χ^2 between validations sources and fit sources for each tract in z-band from the HSC PDR2 dataset, plotted versus the number of visits going into the fit. Blue points show the ratio when the camera model is free to vary. Orange points show the ratio when using a fixed camera model, fit to a large number of visits. Note that the x-axis is cut off at 20 visits, since above this point the difference in the two scenarios becomes negligible.

Second, the camera model can be fed back into the initial single frame processing in the LSST pipeline. Currently, the initial WCS for images comes from combining the telescope pointing with a model for the camera distortions that is hard-coded in the relevant `obs_` package. This model may be theoretically derived or fit from some initial set of observations. This could po-

tentially be improved by using a model generated by the astrometric fitter in the case where the theoretical model does not perfectly capture the real data, or where the camera evolves over time, both very likely scenarios. Again, we can fit the camera model in `gbdesAstrometricFitTask` using some selected set of high-quality data, potentially from a range of pointings. This is then converted into a `lsst.afw.cameraGeom.Camera` object and fed into the single frame processing as a pipeline task input, rather than imported from the `obs_` package. A further advantage of this approach is that changes in the camera over time can be accounted for by splitting observations into separate stable periods, then using exposures from each period to fit a camera model that can be used for the rest of the exposures from that same period.

To construct a `lsst.afw.cameraGeom.Camera` object, `gbdesAstrometricFitTask` should be run with the configuration option `saveCameraObject=True`.

4 Outstanding issues

There are a few known issues that we plan to work on, but which are either in early stages of development or are still in the planning stage:

- **Modeling Atmospheric Turbulence** While exposure times for the HSC precursor data are generally around 300 seconds, the exposure times for LSST will be either 15 or, more likely, 30 seconds. Because of this, atmospheric turbulence is expected to have much stronger effects. Fortino et al. (2021) and Léget et al. (2021) showed that it is possible to model astrometric residuals due to atmospheric turbulence using Gaussian Processes. Initial tests showed that a Gaussian Processes model was able to capture a large portion of the structured astrometric residuals in simulated ComCam data. Figures 9 through 11 show the results of the Gaussian Processes fit for a sample exposure. However, more work is necessary to add this to the pipeline. We will need to confirm that it works well in a larger range of observing conditions on real data, and to ensure that it is not prohibitively expensive in terms of computing resources.
- **Evolution in the Camera** We know that cameras evolve over time, corresponding to various factors such as when the instrument is warmed up and cooled down again. These changes in the instrument lead to changes in the astrometric distortion, which are visible as discontinuities in the distribution of the best-fit visit model parameters. The most straightforward approach to dealing with this will be to run `gbdesAstrometricFit` sepa-

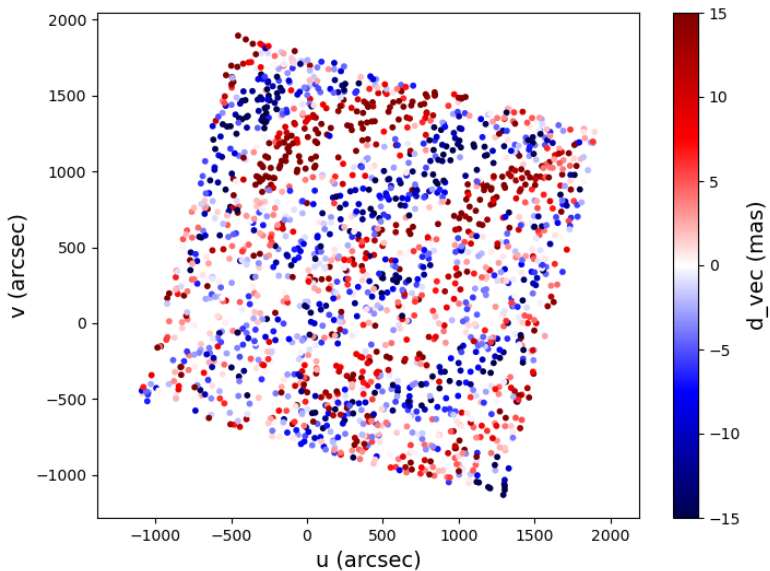


FIGURE 9: Sample astrometric residuals for a simulated ComCam exposure, fit with a low-order polynomial in order to not fit atmospheric turbulence.

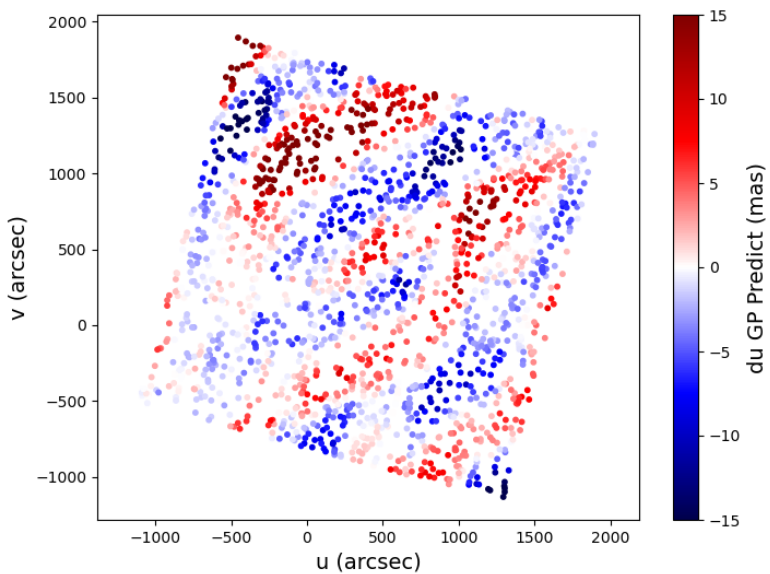


FIGURE 10: Gaussian Processes prediction for the residuals in Figure 9.

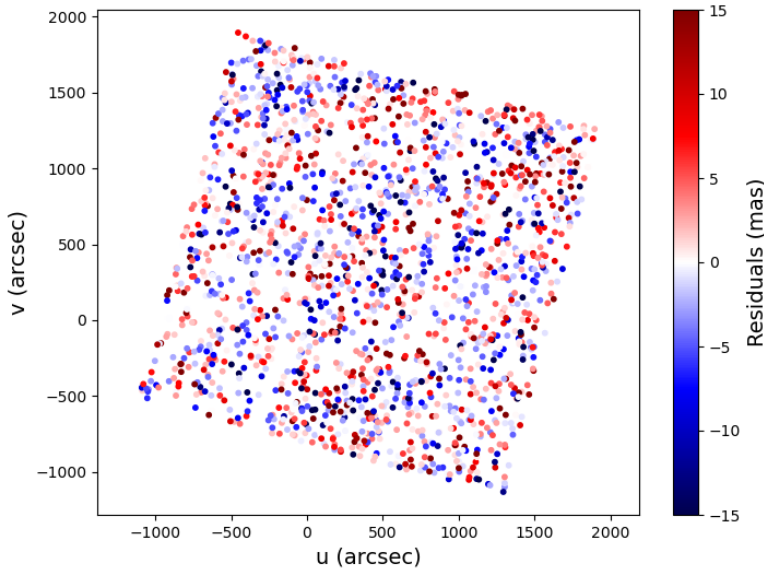


FIGURE 11: Difference between the astrometric residuals and their prediction using Gaussian Processes.

rately for each period of relative stability. A more sophisticated approach might be to allow per-epoch camera models that are fit simultaneously, but it is not clear whether the potential improvement in precision will be worth the added complexity in the fit.

- **Detector Effects** Bernstein et al. (2017) and others have shown that detector electronics and imperfections in the silicon can cause systematic effects in the astrometry, on the scale of a few milliarcseconds for DECam (a similar scale was found for HSC in Léget et al. 2021). These effects, which are expected to be static, can be accounted for by adding an additional mapping to the astrometric model. This mapping can be fit by looking at the mean astrometric residuals from a large number (likely at least 1000) of visits, or can potentially be fit from flats. This mapping would take the form of an additive per-pixel constant vector that would be applied before the per-detector and per-visit maps.
- **The Astrometric Solution for Objects other than Calibration Objects** Five-parameter fits for the coordinates, proper motions, and parallaxes for the bright, isolated point sources that are used in the astrometric fit are a necessary byproduct of the astrometric fit. The fits for these sources will be published as `CalibrationObjects`, possibly with their calibrated fluxes. However, in the long term, we also need to have the same information for the ensemble of objects produced by the pipeline.

In the near to medium-term, we plan to also provide full astrometric fits for all matched sources. This will use the planned, but not yet implemented, table of Object-Source matches, then fit the 5-parameter solution for these objects, keeping the pixel-to-sky coordinate transforms fixed. This will yield the 5-parameter astrometry for those objects that are bright enough to be measured in individual visits. However, these fits are expected to be more affected by blending ambiguity than for the isolated `CalibrationObjects`.

In the longer term, we will provide astrometric fits for all objects, but the method for doing this is still in the planning stages. Potential strategies for doing this include forward-modeling of the per-visit postage-stamps of each object, getting per-visit positions by doing forced centroiding, or measuring centroids in difference-image forced photometry. In some cases, we may want to fit the astrometric solution using object positions in coadded images, stacked either by time period or by season (to isolate parallax). We assume that any of these strategies will require a large number of observations and a time baseline of at least a couple years, which is why this is not considered a priority for development before the survey start.

A References

- Bernstein, G.M., Armstrong, R., Plazas, A.A., et al., 2017, *PASP*, 129, 074503 (arXiv:1703.01679), doi:10.1088/1538-3873/aa6c55, ADS Link
- Fortino, W.F., Bernstein, G.M., Bernardinelli, P.H., et al., 2021, *AJ*, 162, 106 (arXiv:2010.13742), doi:10.3847/1538-3881/ac0722, ADS Link
- Gaia Collaboration, Vallenari, A., Brown, A.G.A., et al., 2023, *A&A*, 674, A1 (arXiv:2208.00211), doi:10.1051/0004-6361/202243940, ADS Link
- Léget, P.F., Astier, P., Regnault, N., et al., 2021, *A&A*, 650, A81 (arXiv:2103.09881), doi:10.1051/0004-6361/202140463, ADS Link
- Shupe, D.L., Moshir, M., Li, J., et al., 2005, In: Shopbell, P., Britton, M., Ebert, R. (eds.) *Astronomical Data Analysis Software and Systems XIV*, vol. 347 of *Astronomical Society of the Pacific Conference Series*, 491, ADS Link

B Acronyms

Acronym	Description
AP	Alert Production
COSMOS	Cosmic Evolution Survey
DCR	Differential Chromatic Refraction
DECam	Dark Energy Camera
DM	Data Management
DMTN	DM Technical Note
DR3	Data Release 3
DRP	Data Release Production
HSC	Hyper Suprime-Cam
ICRS	International Celestial Reference Frame
LPM	LSST Project Management (Document Handle)
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
PDR2	Public Data Release 2 (HSC)
PM	Project Manager
RA	Risk Assessment
WCS	World Coordinate System