Supplementary Material for Pipeline Processing and Data Products for the Roman Galactic Bulge Time Domain Survey and Exoplanet Microlensing Investigation

Roman Science Support Center at IPAC

November 4, 2024

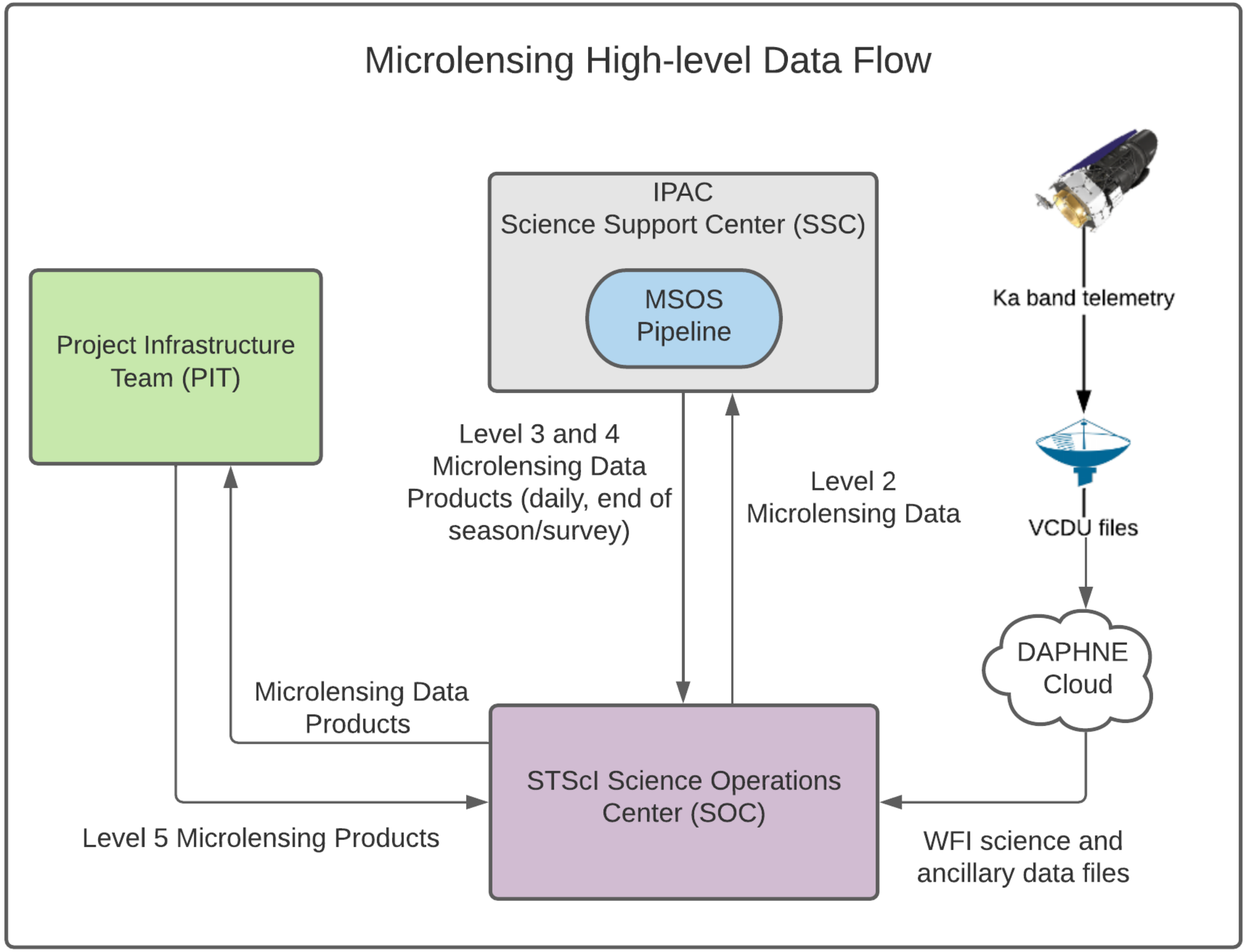
Purpose

- The purpose of these slides is to provide additional details on the processing of data from the Roman Galactic Bulge Time Domain Survey in support of the Exoplanet Microlensing Investigation
- This processing at the Roman Science Support Center (SSC) at IPAC is designed to meet the science goals of the Exoplanet Microlensing Investigation

Roles and Responsibilities

- **SSC: MSOS (Microlensing Science Operations System)**
  - Implement and operate microlensing light curve pipeline
  - Implement and operate event identification and characterization pipeline
  - Implement and operate detection efficiency pipeline
  - Produce data products and send to SOC for archiving
- **SOC**
  - Process all microlensing data through standard WFI Level 1 and 2 pipeline
  - Archive microlensing data products
- **Project Infrastructure Team (PIT)**
  - Scientific analysis of microlensing survey, including planet occurrence rates and other Level 1 science requirements
  - Assist SSC in verification and validation of data products

MSOS Pipeline High-level Data Flow



MSOS Pipeline Outline

- **Introduction**
  - **Roman Galactic Bulge Time Domain Survey**
- **MSOS pipeline**
  - High-level description
    - High-level science flow
    - Temporal modes
    - Pipeline mapping to data products
    - High-level pipeline flow
  - Photometry Pipeline and Light Curve Processing
    - Image Analysis
    - QA
    - Prototype codes and analyses
    - Lens-flux Analysis
  - Event Identification and Characterization
    - Microlensing Event Identification
    - Lensing Model Classification
    - Physical Parameter Determination
  - Detection Efficiency
    - Completeness and Reliability
    - Rhie method analysis

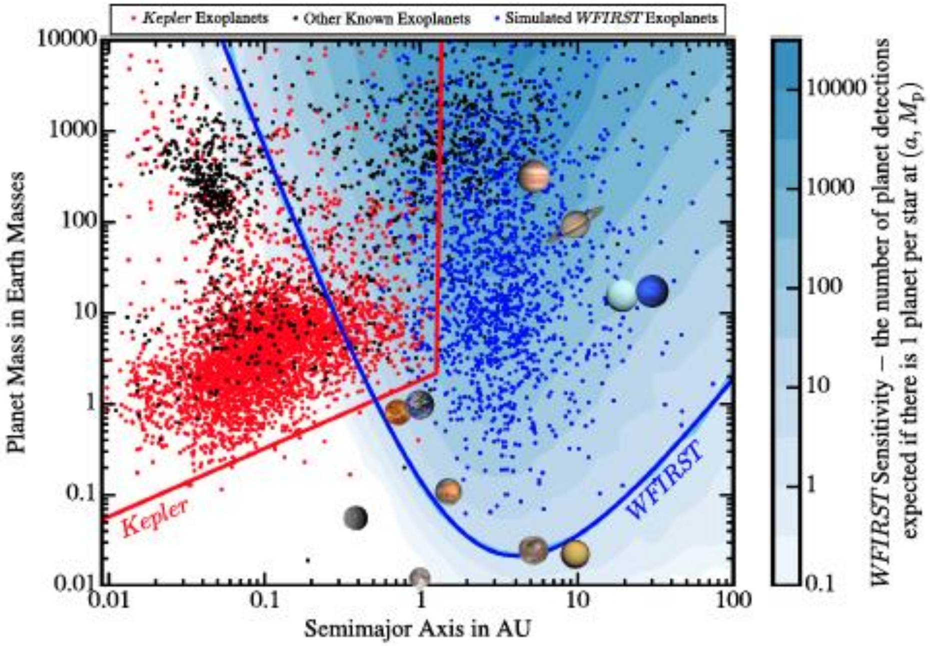
Roman Galactic Bulge Time Domain Survey

GBTDS will be defined by community panel

- Values below used for development but pipeline is robust to number/length of seasons, number of fields etc.

Exoplanet Microlensing Investigation Goal

Statistical census of exoplanetary system from 1 AU to free-floating planets with masses greater than Mars. Expected yield for 1,400 bound exoplanets, including 200 with mass smaller than 3 Earth mass.



Design Reference Mission Observing Strategy

- 7 WFI fields in crowded Galactic Bulge (~ 2 sq deg centered on $l, b \sim 0.8, -1.4$ deg)
- 6 seasons during 5 year primary mission; 62 days per season
- 15 minutes cadence to observe all 7 fields

Raw data products

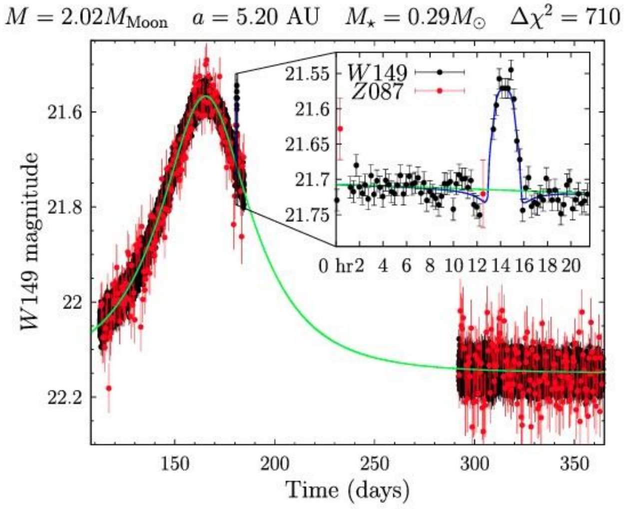
$\sim 41{,}000$ epochs per field (primary filter, $W146$)

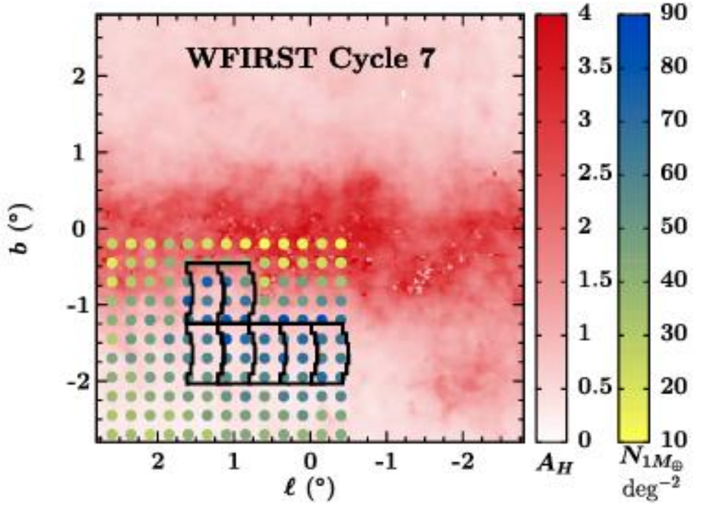
~ 900 epochs per field in two additional filters

$\sim 2 \times 10^8$ monitored stars

Figures (Penny et al. 2019):

- *expected exoplanet sensitivity map in the planet mass vs semi-major axis 2d space*
- *simulated Roman microlensing event (2 seasons) with a planetary anomaly*
- *notional field layout and expected exoplanet microlensing rate (background: H-band extinction, Gonzalez et al. 2012)*

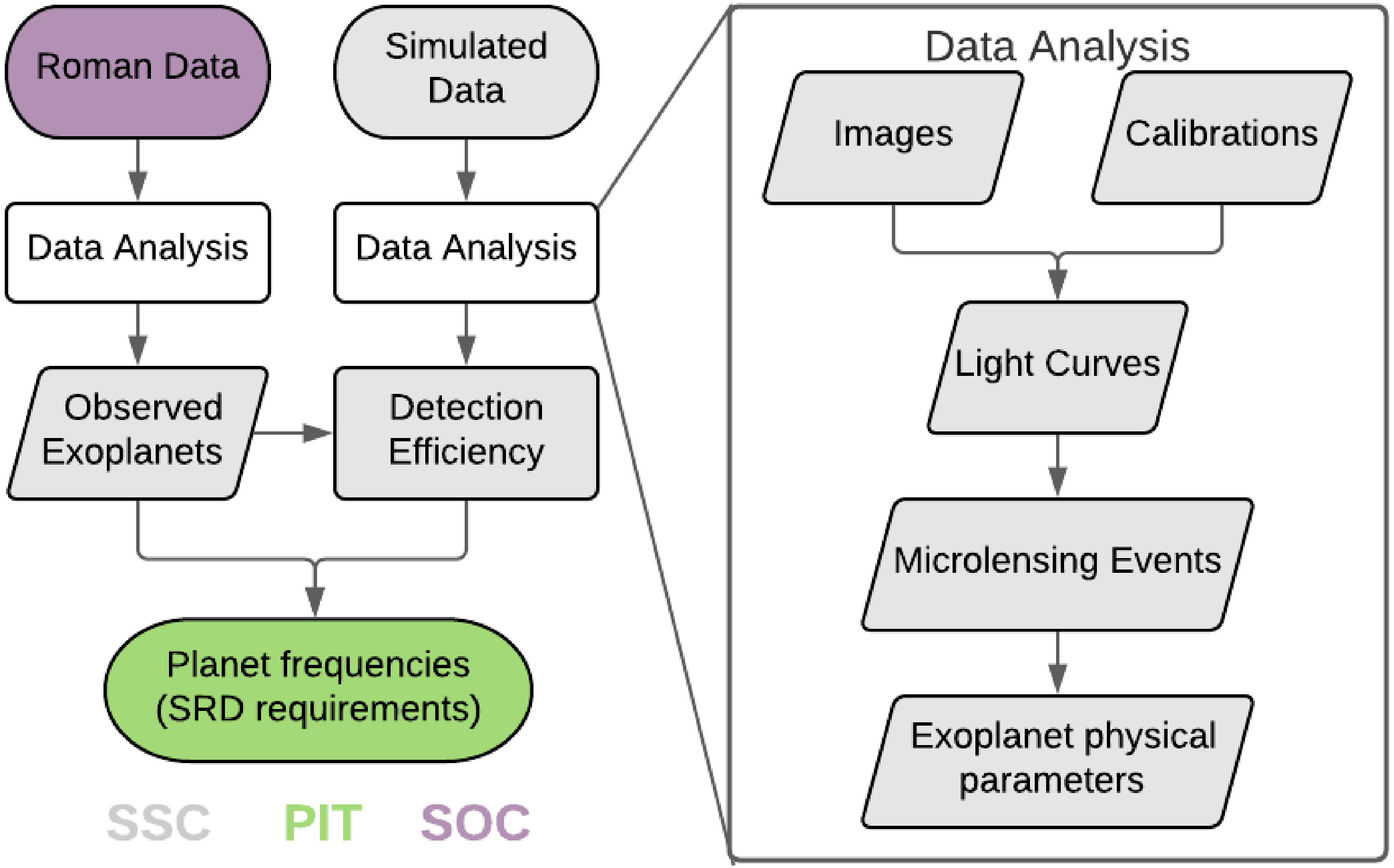




MSOS Pipeline Outline

- Introduction
  - Roman Galactic Bulge Time Domain Survey
- **MSOS pipeline**
  - **High-level description**
    - **High-level science flow**
    - **Temporal modes**
    - **Pipeline mapping to data products**
    - **High-level pipeline flow**
  - Photometry Pipeline and Light Curve Processing
    - Image Analysis
    - QA
    - Prototype codes and analyses
    - Lens-flux Analysis
  - Event Identification and Characterization
    - Microlensing Event Identification
    - Lensing Model Classification
    - Physical Parameter Determination
  - Detection Efficiency
    - Completeness and Reliability
    - Rhie method analysis

MSOS Pipeline High-level Science Flow



MSOS Pipeline Temporal Modes

- **First 30 days**
  - Generate initial reference catalog
  - No data products available during this period but will be produced and made available to archive once complete
- **Daily Pipeline**
  - Produce updates to light curve catalog for ~200 million objects
  - Data made available via archive within 2 days of receipt
  - Every 8 days
    - Update reference image
- **End-of-season pipeline**
  - Recalculate light curve catalog using most recent reference catalog and PSF models
  - Identify variable objects and microlensing events
  - Characterize microlensing events
- **End-of-survey pipeline**
  - Re-process data from all seasons
  - Determine detection efficiency

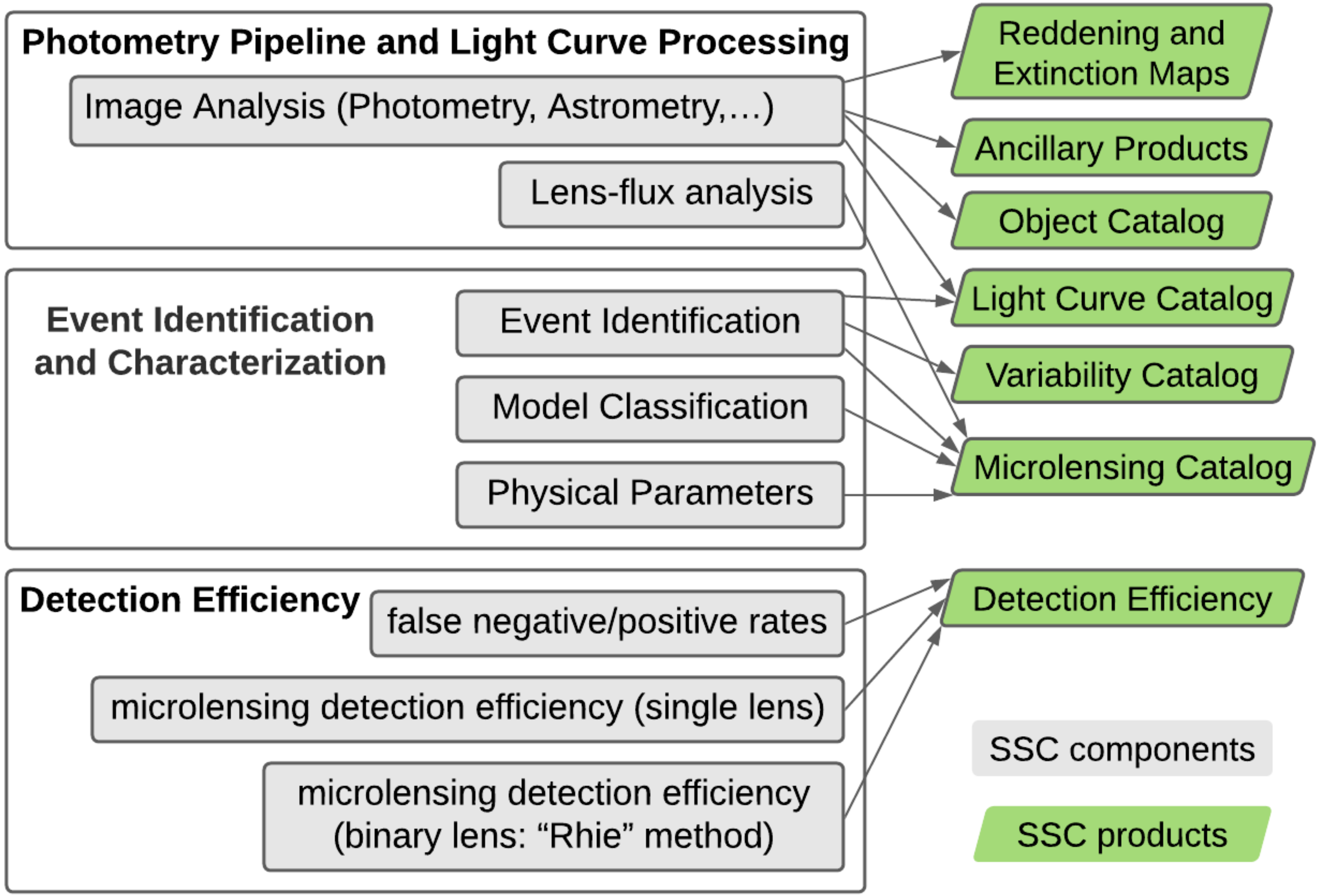
MSOS Data Products

- **Reference frame**
  - wide filter: every 8 days
  - other filters: per season
- **Object Catalog (~200 million sources)**
  - id (linked to Light Curve, Variability and Microlensing Catalogs)
  - position, flux, FWHM: per 8 days
  - proper motion, trigonometric parallax: per season/survey
  - flags: variability, blending, QA
- **Light Curve Catalog for all objects**
  - DIA photometry solution, per exposure
  - PSF photometry solutions, per exposure
  - optimized light curves for microlensing event, season
- **Variability Catalog**
  - Variability metrics
- **Microlensing Event Catalog:**
  - model parameters and QA : single and double lens or source with higher order effects
  - physical parameters of the lens system
  - constraint on the lens flux
- **Detection efficiency**
  - false negative and false positive rates
    - photometry analysis (pixel-level injections)
    - identification and characterization (pixel and catalog-level injections)
  - microlensing detection efficiency
    - single lens: as a function of duration, impact parameter
    - binary lens (Rhie method): microlensing (q, s) and physical (mass, separation) parameter spaces, per event and average
- **Reddening and extinction maps**

Level-3

Level-4

MSOS Pipeline mapping to data products



MSOS Pipeline Outline

- Introduction
  - Roman Galactic Bulge Time Domain Survey
- **MSOS pipeline**
  - High-level description
    - High-level science flow
    - Temporal modes
    - Pipeline mapping to data products
    - High-level pipeline flow
  - **Photometry Pipeline and Light Curve Processing**
    - **Image Analysis**
    - **QA**
    - **Prototype codes and analyses**
    - **Lens-flux Analysis**
  - Event Identification and Characterization
    - Microlensing Event Identification
    - Lensing Model Classification
    - Physical Parameter Determination
  - Detection Efficiency
    - Completeness and Reliability
    - Rhie method analysis

Photometry Pipeline and Light Curve Processing
Image Analysis

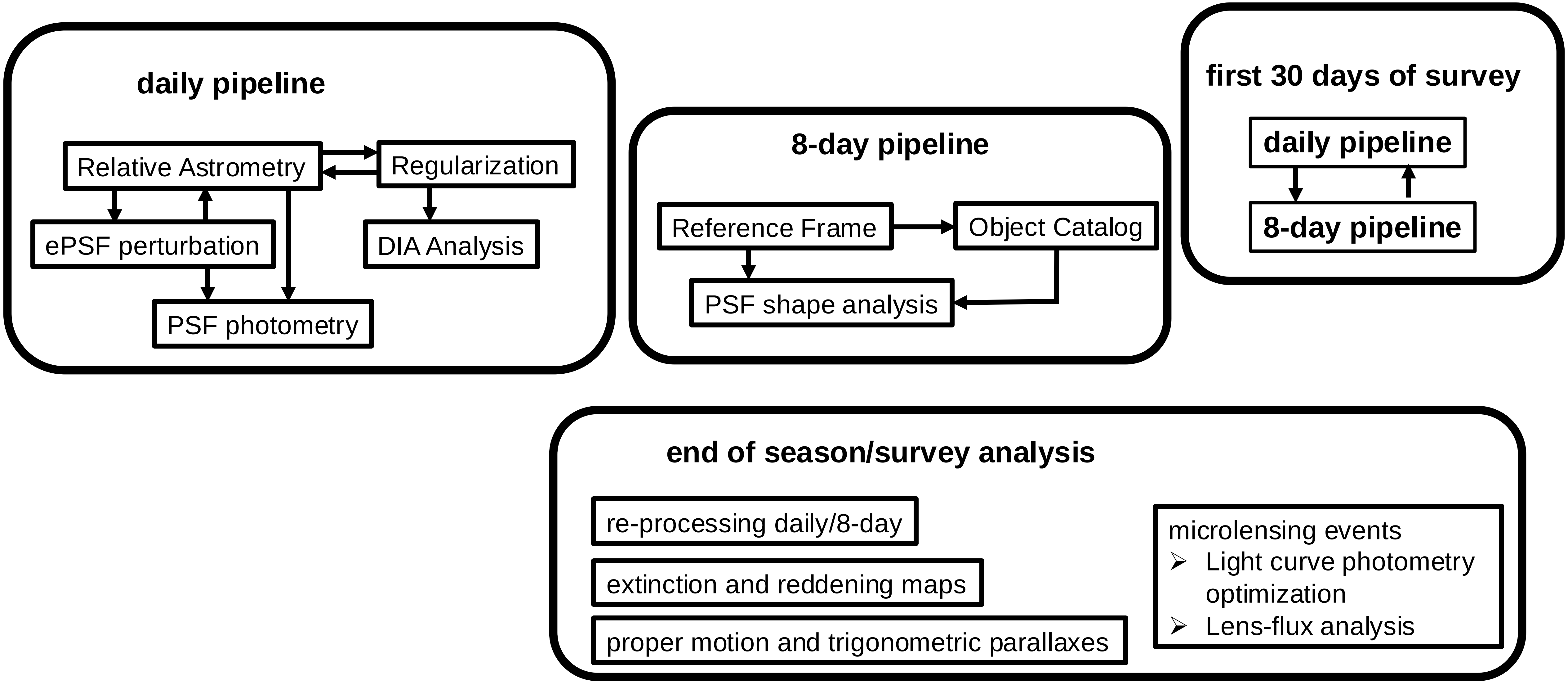


Image Analysis (1)

- The Reference Frame is a stacked realization of the astrophysical scene built out of 8-days worth of data, given their corresponding astrometry solution
- The Roman/WFI detector is undersampled
- "effective" PSF (ePSF) procedure introduced for the analysis of undersampled data (Anderson and King 2000) and used thereafter in particular for the analysis of HST and Spitzer/IRAC data
- The Reference Frame is effective and oversampled in the same sense as the ePSF
- The mechanism for evaluating the "residual exposure", the blotted Reference Frame minus a single exposure, is at the basis of the Difference Image Analysis (DIA)

Reference Frame module flow chart

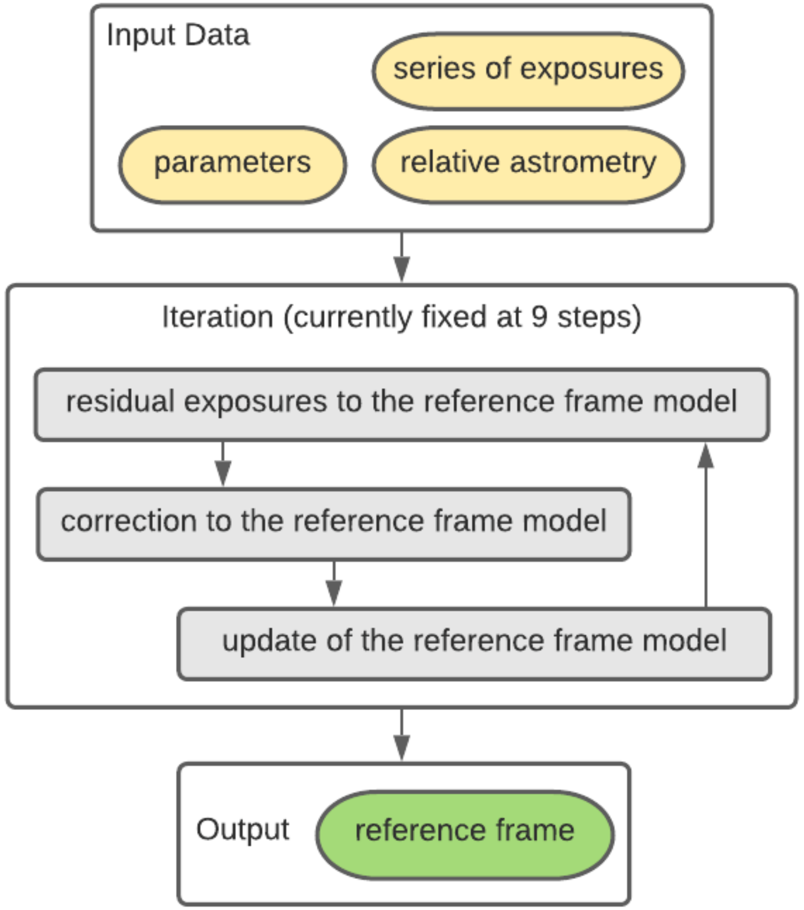


Image Analysis (2)

- **Relative astrometry (intermediate product)**
  - linear transformation versus Reference Frame using input detector Geometric Distortion Solution and exposure World Coordinate Systems
- **ePSF evaluation (intermediate product)**
  - "effective" PSF following procedure from Anderson & King (2000) - for undersampled detectors
  - ePSF for the Reference Frame with variations for each exposure handled perturbatively
- **Reference Frame**
  - noiseless effective and oversampled (along the same lines of the ePSF) stacked (8-days of data for wide filter) realization of the astrophysical scene, input for the following:
    - Object Catalog
    - DIA analyses
    - PSF shape evaluation
    - lens-flux analysis (end of season/survey, microlensing events only)
- **Object Catalog (based upon the Reference Frame, updated every 8 days)**
  - identification and photometry for the ensemble of the monitored objects

Image Analysis (3)

- **Regularization (intermediate product)**
  - procedure to bring each given exposure to the same ePSF model of the Reference Frame (a technical point: the procedure does not involve the use of the ePSF)
    - regularized exposures, input for the Reference Frame
    - regularized residual exposure, input for the DIA analyses
- **DIA analyses**
  - differential DIA photometry for the ensemble of the Object Catalog
  - search for new transients
- **PSF Photometry**
  - analysis carried out on the input (Level 2) exposure with adapted ePSF model per exposure
  - procedure for each catalog object (the “target”)
    - remove all known neighbor stars
      - PSF-fitting for the flux with fixed position of the target star
      - PSF-fitting for both flux and position of the target star
    - PSF-fitting for target and neighbor stars with fixed position

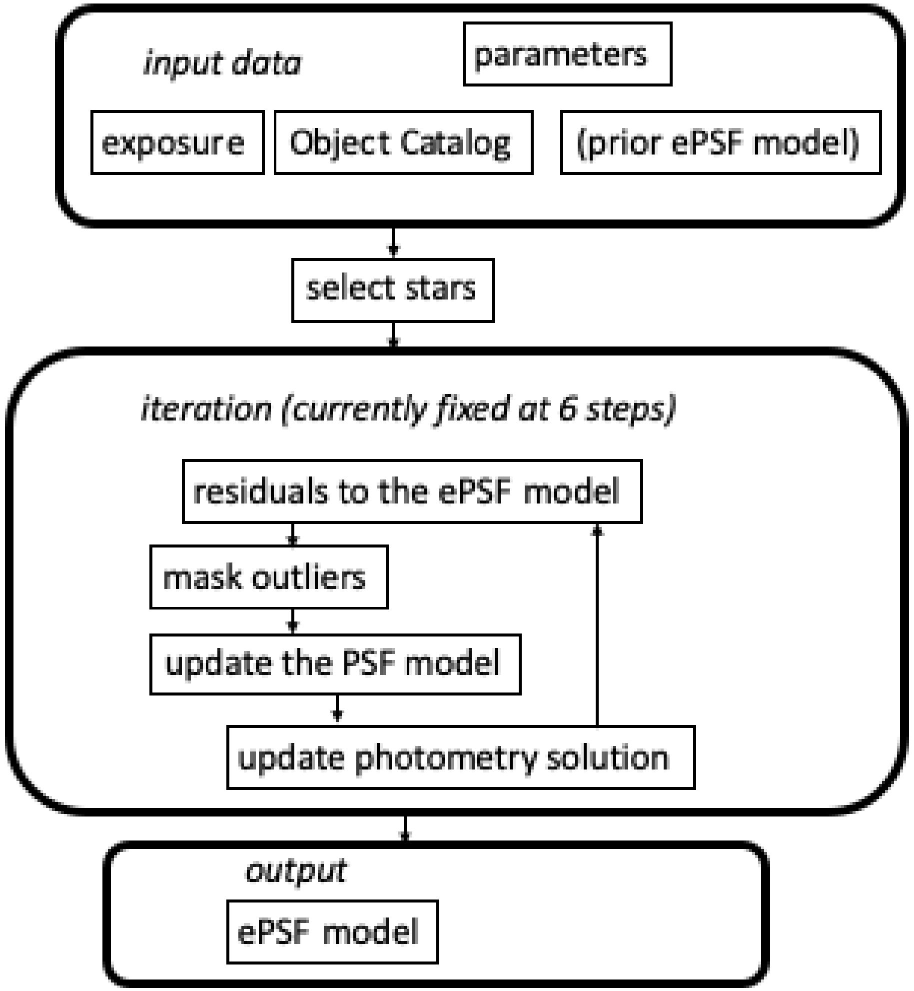
Image Analysis (4)

- **Light Curve Photometry Optimization (end of season/survey, microlensing events only)**
  - DIA: optimization Reference Frame and centroid (given microlensing model)
  - PSF: optimization ePSF model, optimization neighbor stars and simultaneous two-star fit
- **PSF shape analysis (8-days): FWHM measurement based upon residual of the Reference Frame after subtraction of neighbor stars: PSF-fitting after convolution with a pre-defined set of gaussian**
- **Proper motion and trigonometric parallax (end of season/survey): least square minimization given the astrometry as a function of time and the corresponding spacecraft position (standard procedure following eg Kirkpatrick et al. 2019)**
- **Reddening and Extinction Maps (intermediate product, needed for microlensing event characterization analysis)**
  - standard procedure based, eg, on Nataf et al. (2013), Nataf et al. (2016)
    - characterization of the Red Clump: magnitude and color (all the available filters)
    - reddening: the difference of the observed from the reference Red Clump color
    - extinction: linear regression red clump magnitude versus reddening

Image Analysis: QA

- **Data quality assessment for each product**
- **Specific QA Level 5 requirements to be tested on the data**
  - relative photometric statistical precision per exposure
  - relative photometric systematic precision over a season
  - relative photometric systematic precision across seasons
  - relative astrometric statistical precision per exposure
  - relative astrometric systematic precision over a season
  - photometry absolute calibration
  - photometry relative zero precision
  - FWHM precision, per measurement
  - FWHM precision, along the survey

Image Analysis: ePSF model



analysis for undersampled detectors with effective PSF (ePSF) procedure Anderson & King (2000)

$$P_{ij} = s_* + f_* \, \psi_E(i - x_*, j - y_*)$$

$$\hat{\psi}_E(i - x_*, j - y_*) = \frac{P_{ij} - s_*}{f_*}$$

key parameters:

- star selection
- oversampling factor
- spatial extension

note: the prior ePSF model may be identically null

the use of the ePSF is ubiquitous throughout the photometry analysis

Image Analysis: Object Catalog

- ❑ search and characterization of point-like objects carried out on the Reference Frame given the corresponding ePSF model

- ❑ iterative analysis:
  - residual image subtracting current model
  - *(slowly)* identify new objects as local maxima
  - update catalog photometry solution

- ❑ key parameters
  - number of iterations
  - sky determination
  - object/star selection threshold values
    - isolation
    - signal-to-noise
    - “slow down” thresholds for nearby stars identification
  - photometry (ePSF)
    - maximum number nearby stars fitted simultaneously (baseline plan: 2)

Object Catalog flow chart

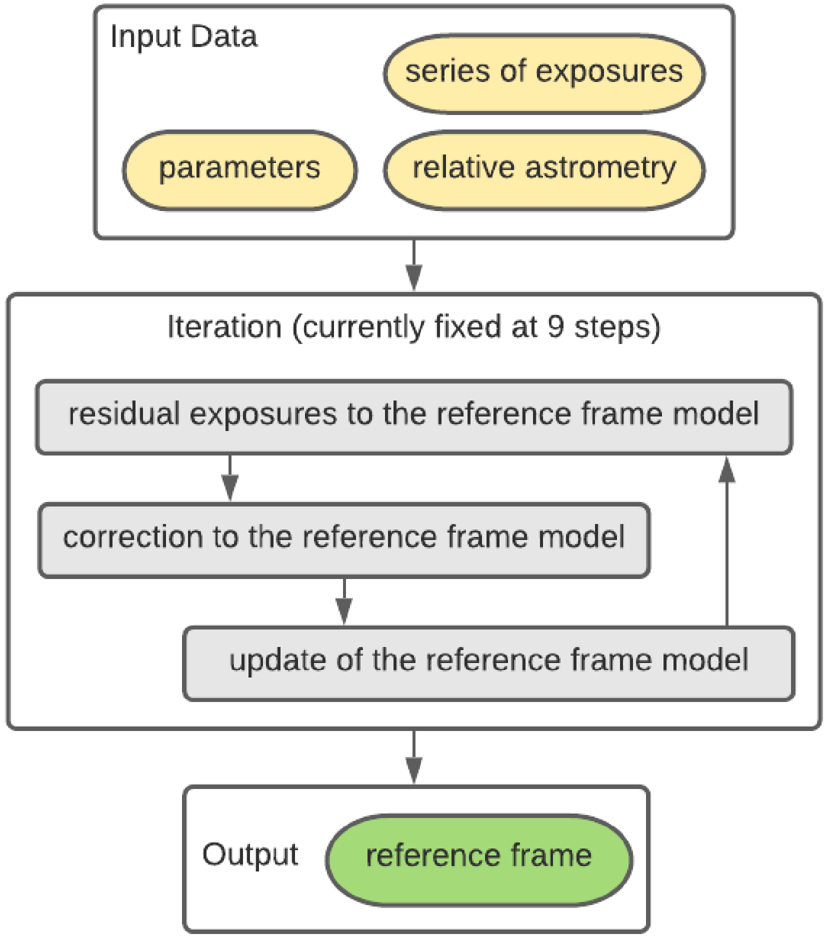
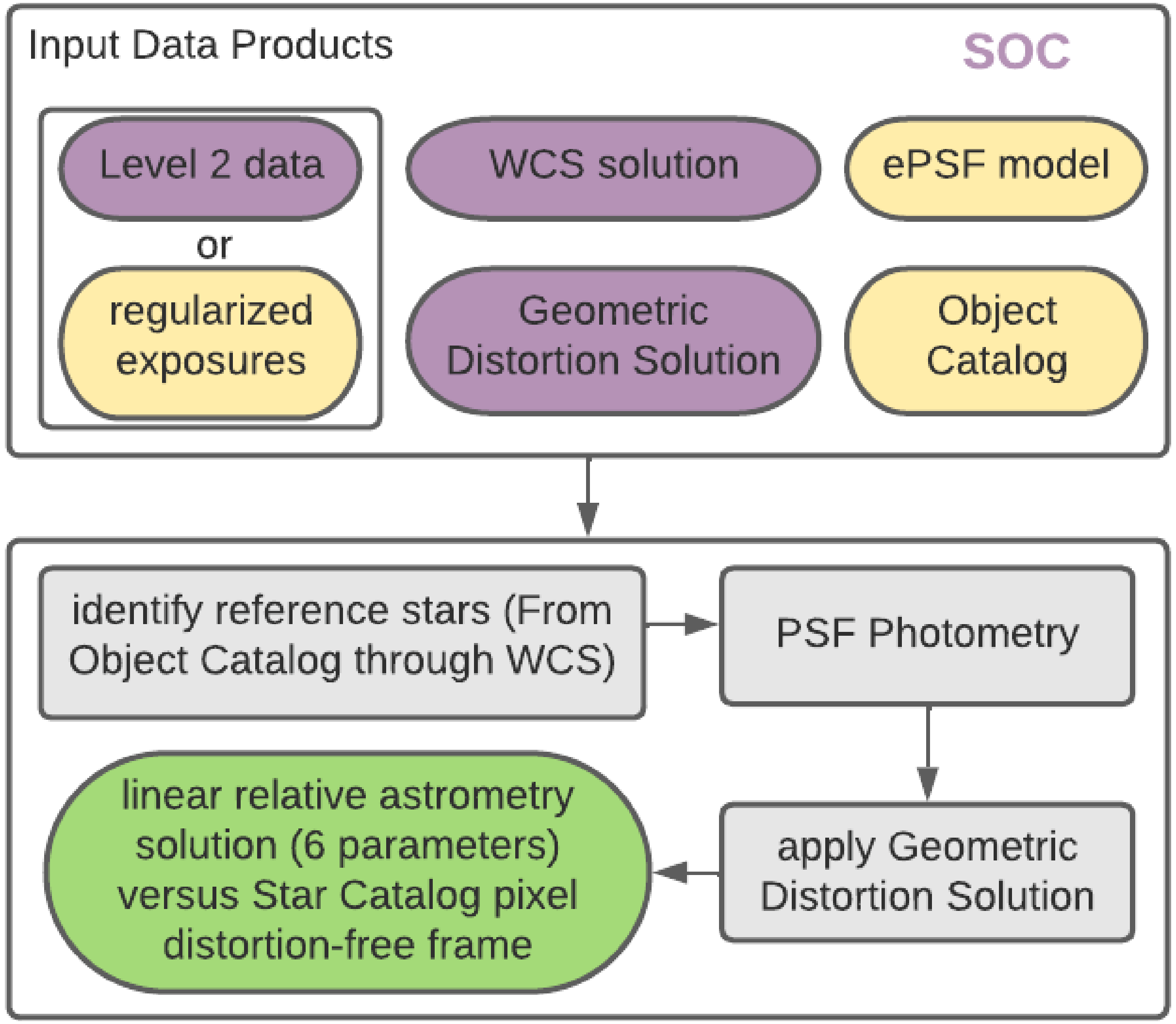


Image Analysis: Relative Astrometry



Object Catalog initialization

- distortion-free pixel coordinate system
- pixel to Galactic coordinates(eg, GAIA) transformation

Relative astrometry solution per exposure versus the Object Catalog pixel coordinate system

first solution for Level 2 data:

- ➢ average ePSF model

iterated solution for Level 2 data:

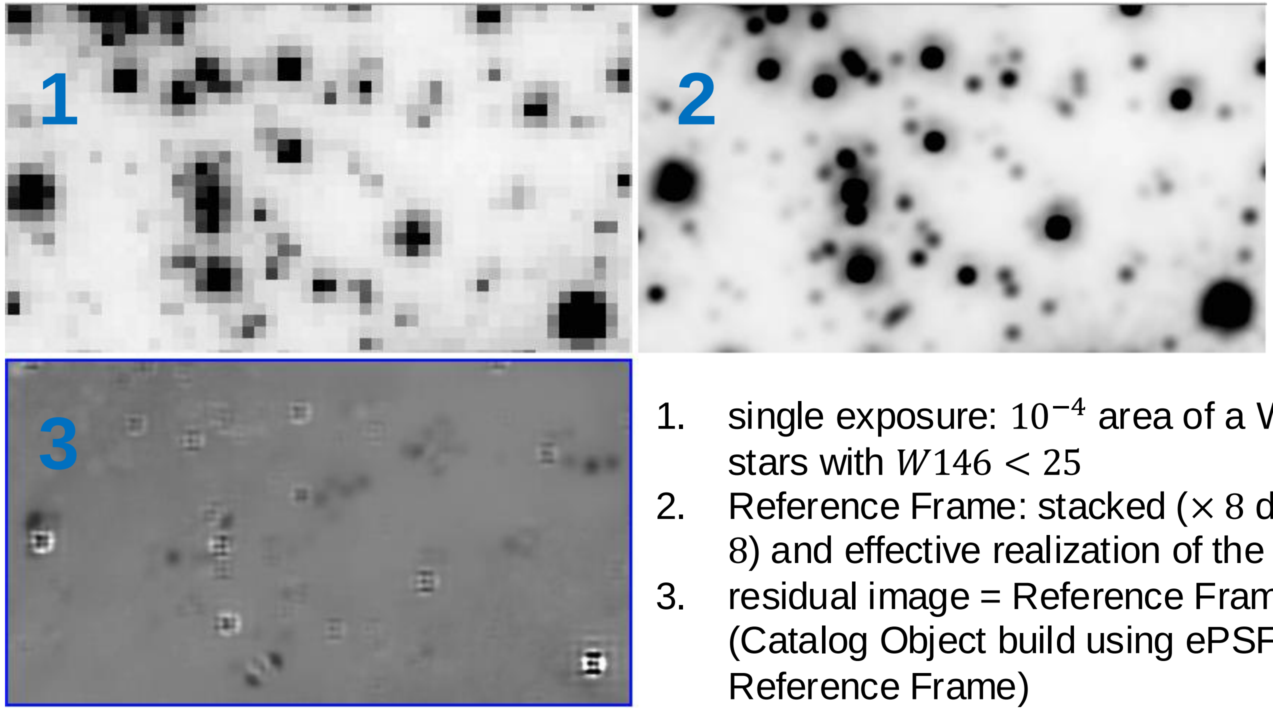
- ➢ perturbed ePSF model

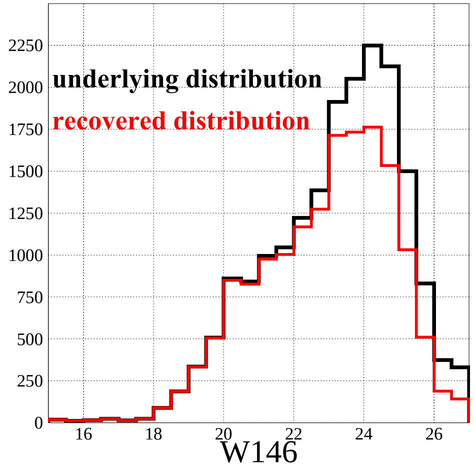
solution for regularized Level 2 data:

- ➢ average ePSF model

Prototype codes and analyses

- **Analysis based upon a point source Roman/WFI simulation developed within the SIT**
  - astrophysical scene and detector layout from Penny et al. (2019)
  - PSF model from Roman adaption of STScI/WebbPSF (Perrin et al. 2014)
- **Simulation for a series of 750 dithered exposures for the Wide filter (as for 8-day pipeline)**
- **Prototype codes for the analysis by Jay Anderson (SIT)**
  - Reference Frame
  - Object Catalog
  - Regularization
  - ePSF extraction





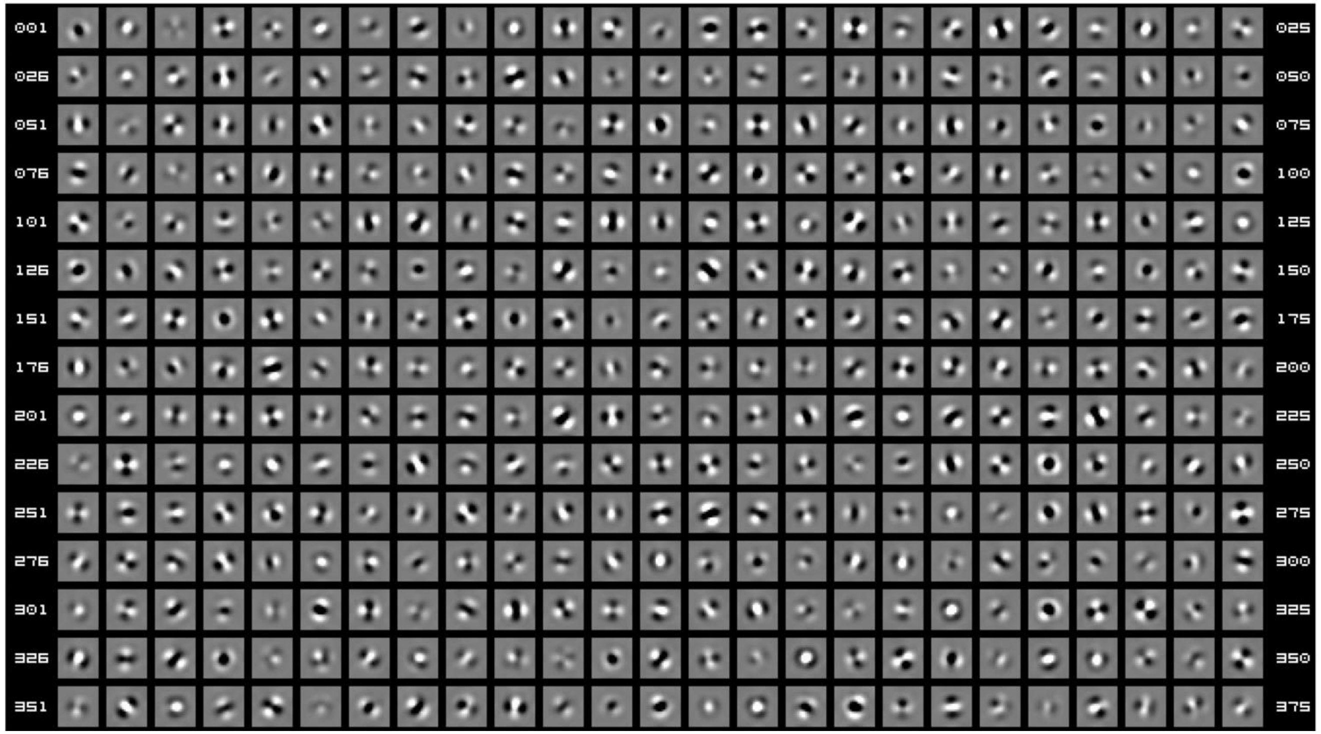
1. single exposure: 10^{-4} area of a WFI detector, includes ~ 200 stars with $W146 < 25$
2. Reference Frame: stacked ($\times$ 8 days of data), oversampled ($\times$ 8) and effective realization of the astrophysical scene
3. residual image = Reference Frame – model Catalog Object (Catalog Object build using ePSF model extracted from the Reference Frame)

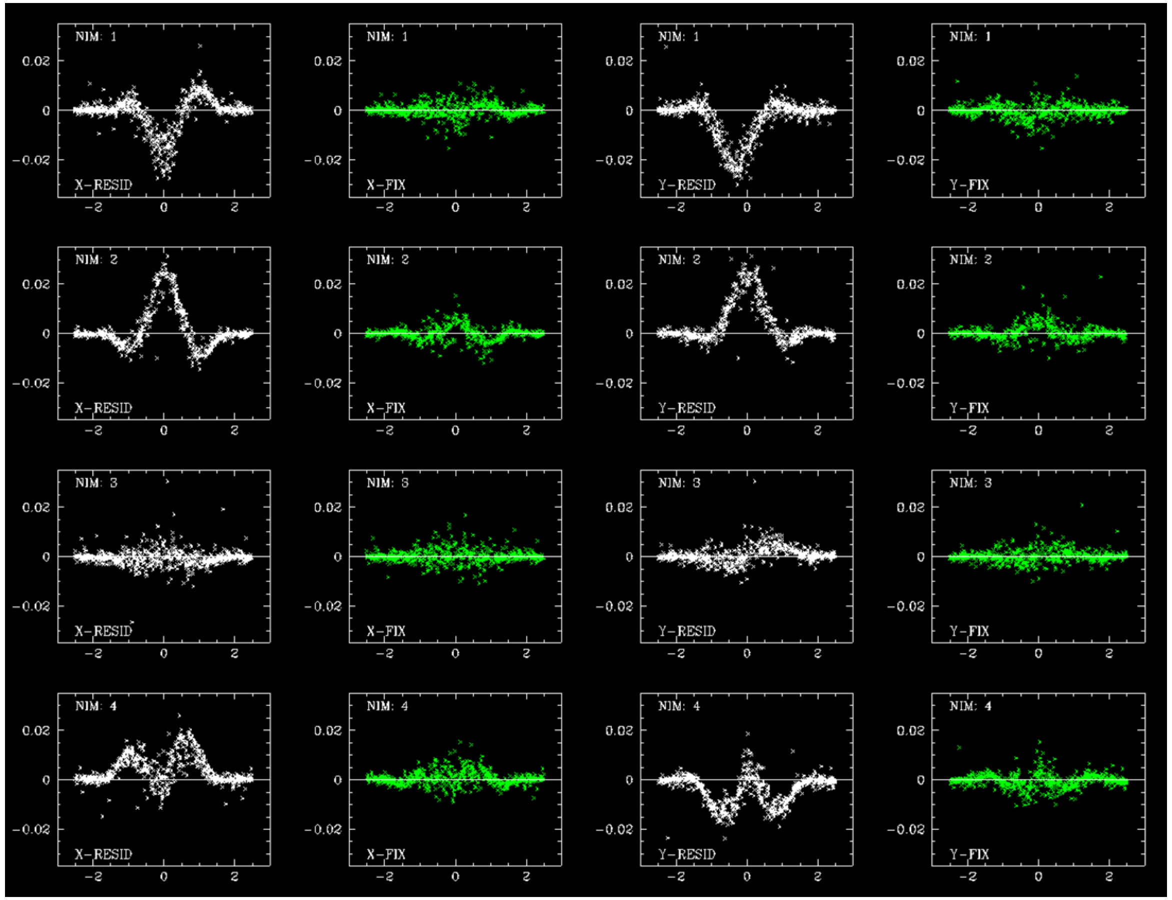
Image Analysis
Prototype codes and analyses

Regularization

Analysis for a series of simulated exposures with jitter

- PSFs residual before (white) and after regularization
- Jitter pattern is different in each exposure vs average



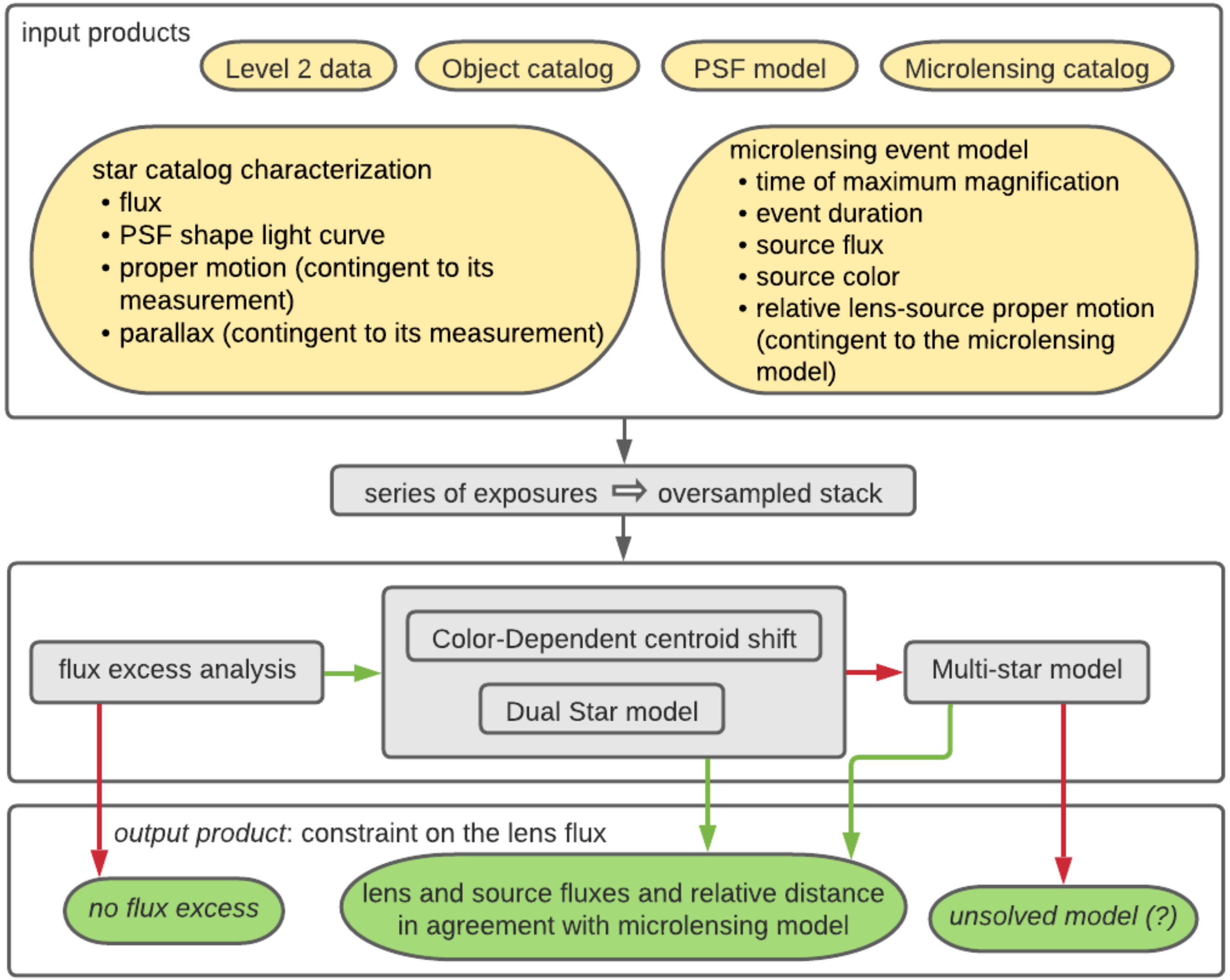


Figures from Jay Anderson

Photometry Pipeline and Light Curve Processing

Lens-flux Analysis

- Refined photometry analysis (at the image level), restricted to the selected sample of microlensing events, to constrain the lens flux using information from both the MOSOS Object Catalog and the underlying microlensing model
- The expected maximum lens-source separation along the 5 years of the survey is of about 0.3-0.4 pixels
- The photometry analysis is carried out on stacked images built analogously to the Reference Frame, for all available filters
- Analysis procedure as outlined in Bennett, Anderson, Gaudi (2007)



MSOS Pipeline Agenda

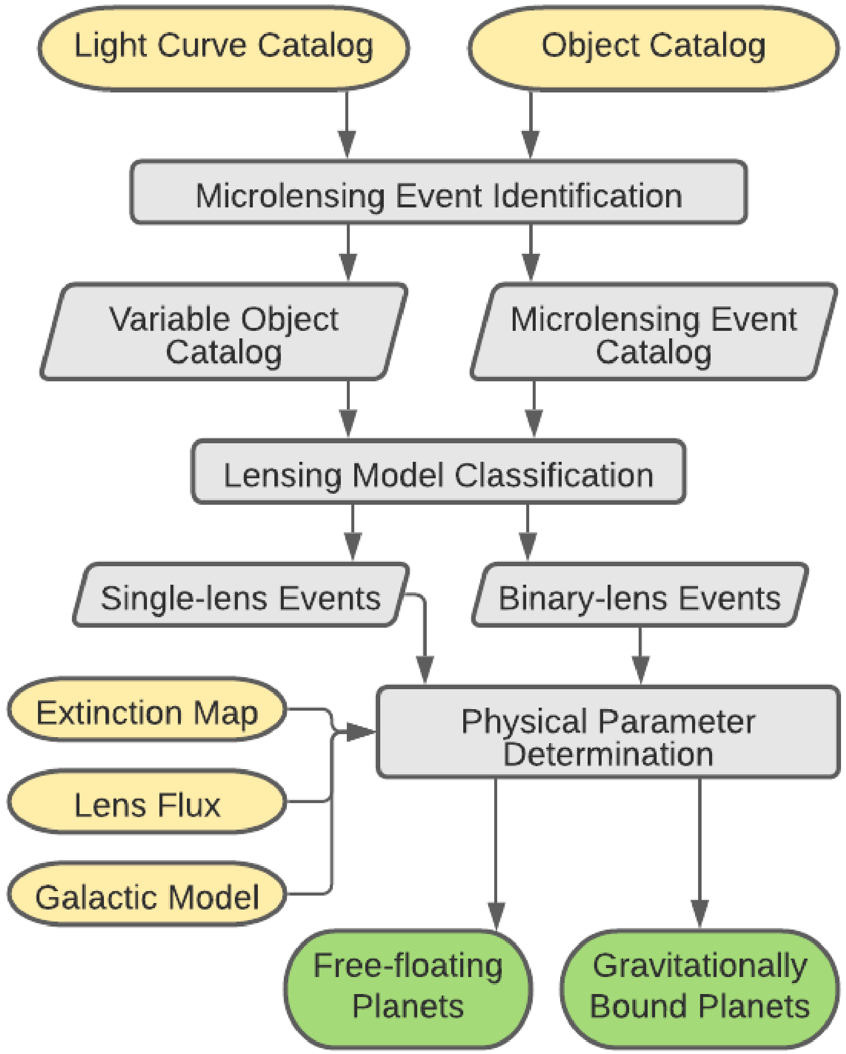
- Introduction
  - Roman Galactic Bulge Time Domain Survey
- **MSOS pipeline**
  - High-level description
    - High-level science flow
    - Temporal modes
    - Pipeline mapping to data products
    - High-level pipeline flow
  - Photometry Pipeline and Light Curve Processing
    - Image Analysis
    - QA
    - Prototype codes and analyses
    - Lens-flux Analysis
  - **Event Identification and Characterization**
    - **Microlensing Event Identification**
    - **Lensing Model Classification**
    - **Physical Parameter Determination**
  - Detection Efficiency
    - Completeness and Reliability
    - Rhie method analysis

Event Analysis: Overview

- **Three Components**
  - Microlensing Event Identification
  - Lensing Model Classification
  - Physical Parameter Determination

- **Timescales:**
  - End-of-season
    - Detect new events
    - Refine model(s) for known events
  - End-of-survey
    - From “scratch”

- **Default values selected for all parameters**
  - Can be adjusted if needed after testing or validation

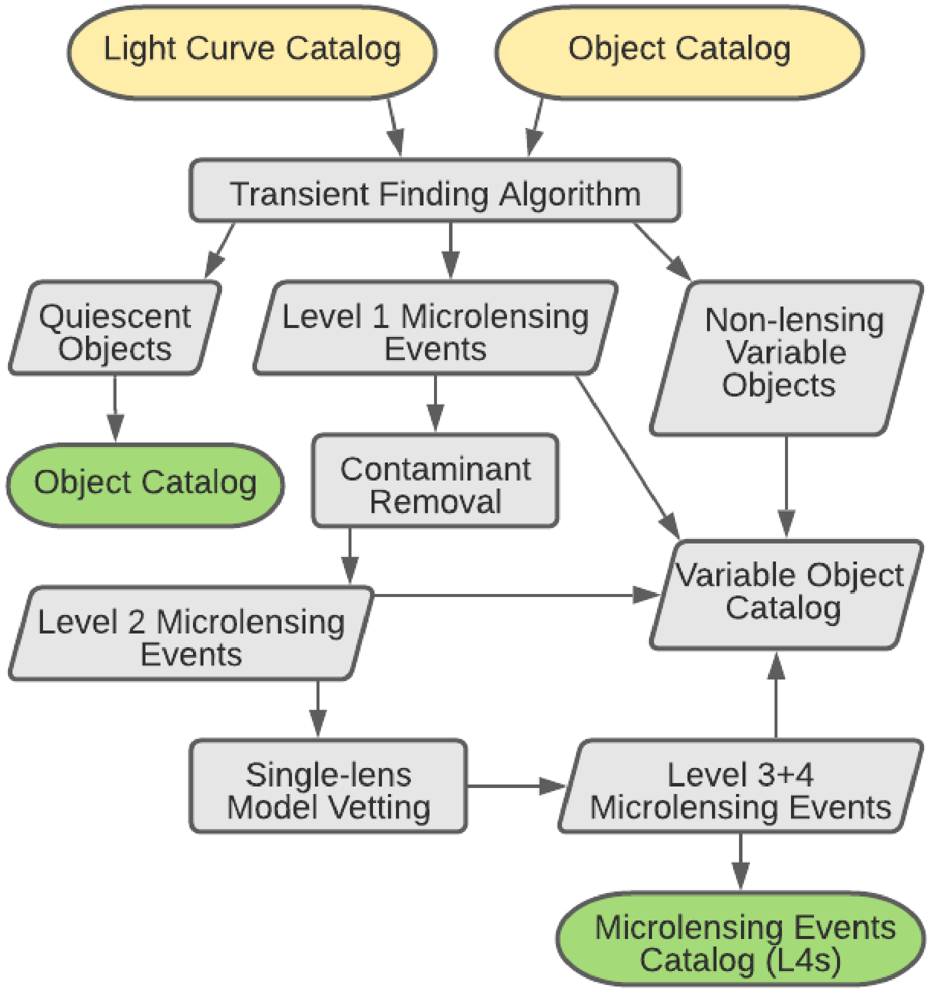


Event Analysis I: Microlensing Event Identification

- **Based on procedures applied to ground-based microlensing surveys**
  - Identify and characterize photometric brightenings
  - Identify, characterize, and flag contaminants
    - Chromaticity
    - Brightening concurrence
    - Low-amplitude variability
  - Evaluate lensing morphology congruence
  - Community packages exist for several areas

- **Input Products**
  - Light Curve Catalog
  - Object Catalog

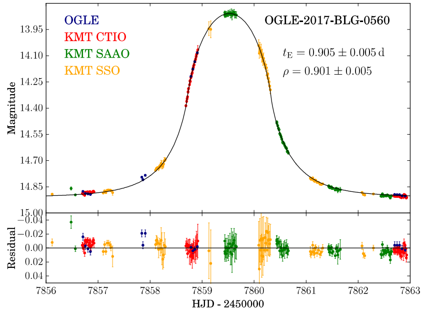
- **Output Products**
  - Object Catalog (updated)
  - Variable Object Catalog
  - Microlensing Event Catalog

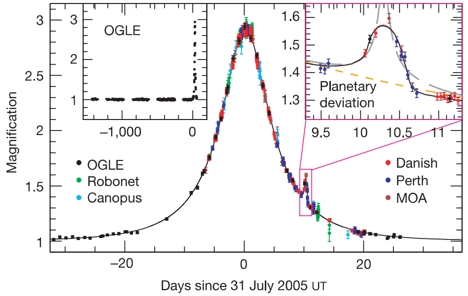


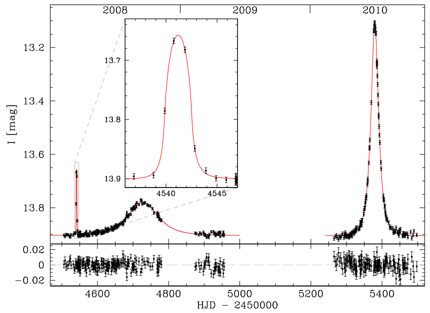
Event Analysis I: Microlensing Event Identification

Single Brightening

Multiple Brightenings

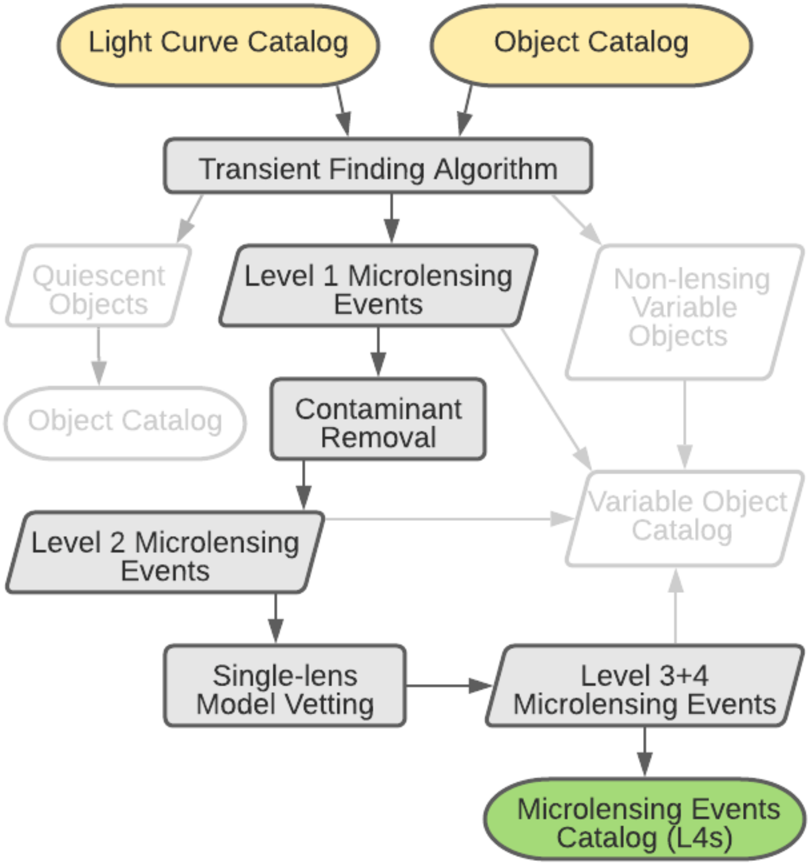






Single-lens Free-floating Planet Candidate

Binary-lens Gravitationally Bound Planet



Beaulieu+ (2006); Mróz+ (2019); Poleski+ (2014)

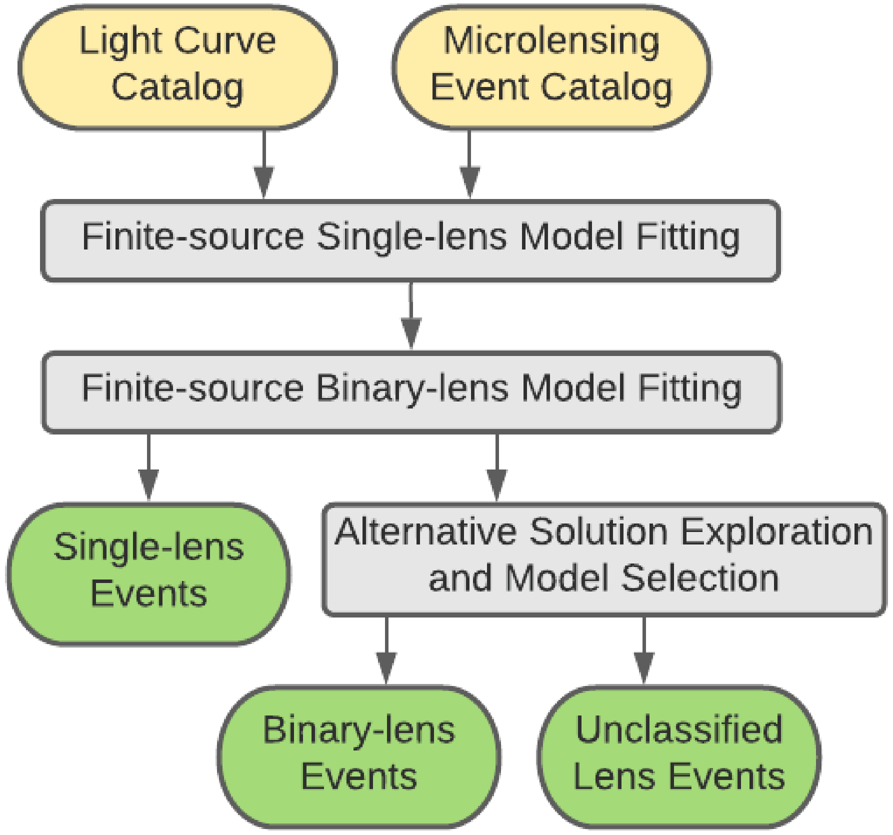
cf. Mróz+ (2019,2017); Sumi+ (2011); Kim+ (2018a)

Event Analysis II: Lensing Model Classification

- **Light Curve Magnification Computation Algorithms:**
  - `VBBL`: Publicly available (Bozza+ 2020,2018; Bozza 2010)
  - `ICRS`: Provided by MicroSIT (Bennett 2010; Bennett & Rhie 1996)

- **Posterior Distribution Sampling:**
  - `sfit`: Provided by MicroSIT
  - `EMCEE`: Publicly available (Foreman-Mackey+ 2013; Goodman & Weare 2010)

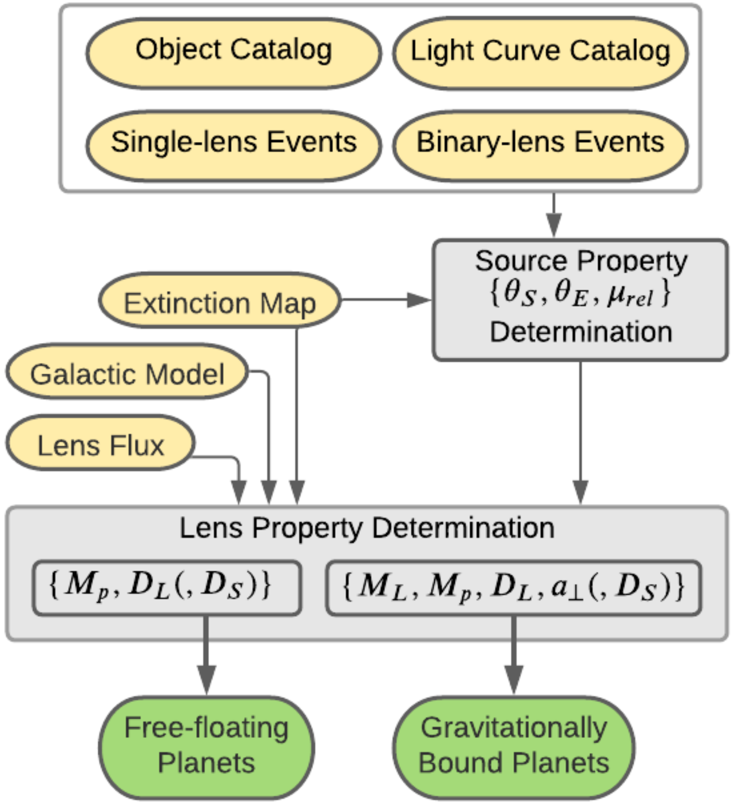
- **Tiered Model Calculation and Selection**
  - Finite-source single-lens modeling
  - Finite-source binary-lens modeling
    - Peak identification
    - Template fitting
    - Higher-order effects
  - Alternative solution exploration
  - MCMC analysis for best-fit models



Event Analysis III: Physical Parameter Determination

- **Properties of ‘Source’ System (i.e., the background star)**
  - Follow procedure of Yoo+ (2004)
  - Derive de-reddened photometric characteristics
  - Determine filter-specific (linear) limb-darkening coefficients
    - Re-run MCMC modeling
  - Constrain physical properties

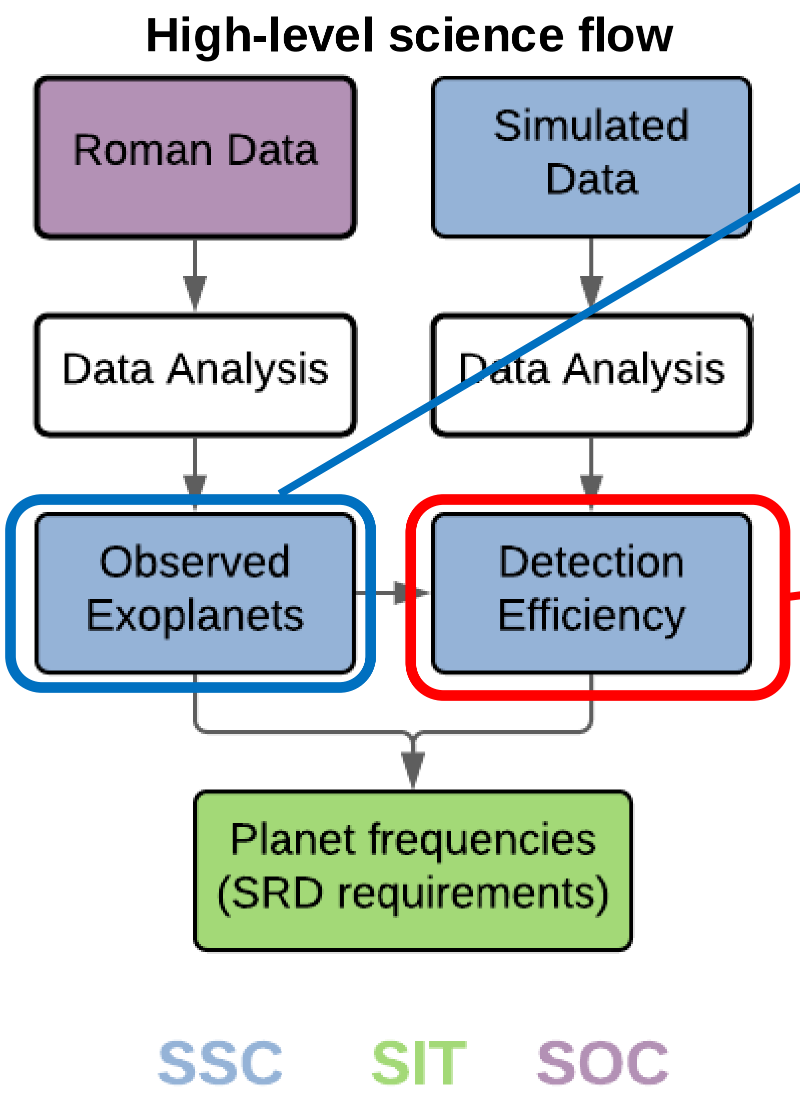
- **Properties of ‘Lens’ System (i.e., the planetary system)**
  - Planet and host star masses
    - If higher-order effects measured, compute directly
    - Else, apply Bayesian prior from Galactic model
  - Source distance
    - If trigonometric parallax measured, compute directly
    - Else, apply Bayesian prior from Galactic model
  - Lens distance and orbital semi-major axis
    - If higher-order effects measured, compute directly
    - Else, apply Bayesian prior from Galactic model
  - Adopt Galactic model of Poleski+ 2020 (based on Clanton & Gaudi 2014)



MSOS Pipeline Agenda

- Introduction
  - Roman Galactic Bulge Time Domain Survey
- **MSOS pipeline**
  - High-level description
    - High-level science flow
    - Temporal modes
    - Pipeline mapping to data products
    - High-level pipeline flow
  - Photometry Pipeline and Light Curve Processing
    - Image Analysis
    - QA
    - Prototype codes and analyses
    - Lens-flux Analysis
  - Event Identification and Characterization
    - Microlensing Event Identification
    - Lensing Model Classification
    - Physical Parameter Determination
  - **Detection Efficiency**
    - **Completeness and Reliability**
    - **Rhie method analysis**

Detection Efficiency: Rationale and Procedure

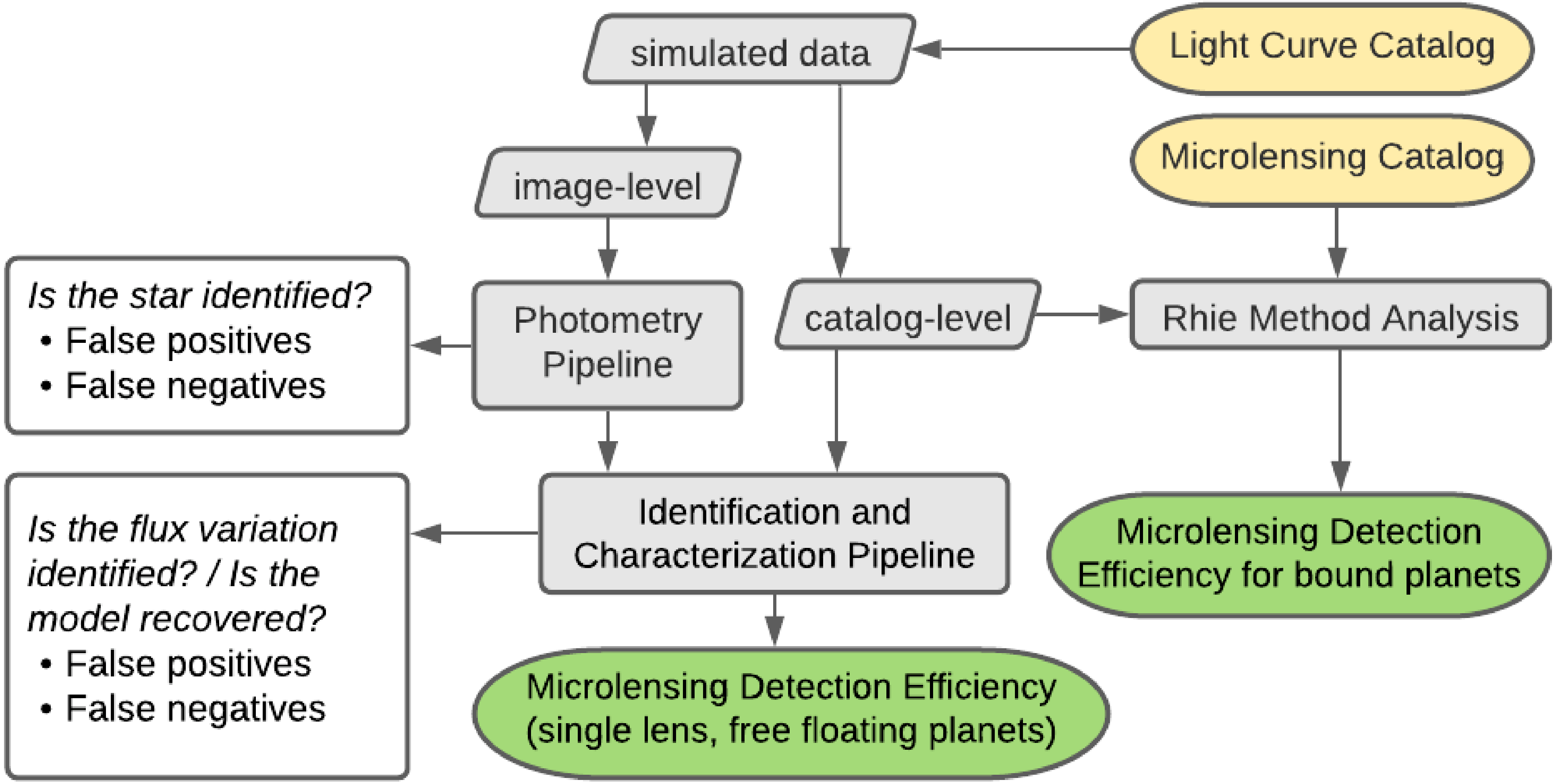


- The data analysis over Roman data provides a sample of (microlensing) exoplanets and corresponding microlensing and physical parameters

- The detection efficiency is a necessary ingredient to address the occurrence rate issues, and in particular the Planet Frequencies SRD requirements

- The goal of the detection efficiency is to assess how well the pipeline is able to recover the signal: to identify and characterize microlensing events and distinguish them from other flux variations

Detection Efficiency: Processing Flow



Detection Efficiency
Rhie Method Analysis

- **Rationale of the method**
  - assess whether a given (binary) planetary model can be detected from the data
- **Input products**
  - Light Curve Catalog
  - Microlensing Catalog (the ensemble of detected events searched for exoplanets)
  - Astrophysical prior (grid of values to be tested for binary microlensing magnification model)
- **Output products**
  - Detection Efficiency as a function of the binary microlensing parameters (q, s): per microlensing event and average
  - Detection Efficiency as a function of planet mass and separation (physical units): per microlensing event and average
- **Procedure**
  - catalog-level injection of a simulated binary lens model (with the underlying single-lens model parameters as in observed events), provided a noise model
  - single lens model least square minimization to assess whether the planet can be detected
  - efficiency is the fraction of detected events after marginalization over non-essential parameters
  - efficiency in physical parameter space based upon prior on lens mass and distance
- **Key parameters**
  - grid in the binary microlensing space and for finite source effect
  - threshold $\Delta\chi^2$ for detection

Roman Data Levels

Level 0	Science telemetry: Packetized data as it arrives from the spacecraft
Level 1	Uncalibrated exposures: include metadata, engineering data
Level 2	Calibrated exposures: remove detector signature, map to scene flux
Level 3	Resampled data: use rectified grid, coadd multiple exposures as appropriate
Level 4	High-level data: mostly source-oriented, include catalogs, extracted spectra, postage stamps, high fidelity flux, shape, morphology measurements
Level 5	Community-contributed products: may include any of the above types

35