Hyper Suprime-Cam Catalogue Columns and Flag Filters

This document provides a brief outline of the necessary columns to extract from HSC to create the smaller dataset used for cross-matching. It also provides more information on the “flag” index, additional columns to extract and combine into a quality flag value.

Each catalogue will need to be transformed from its current state in the DAC to separate files for ingestion into the cross-match process, to allow for a common interface for flexible matching across multiple different catalogues. A summary of this in its current form is provided below.

NOTE: HSC is split into multiple databases width-wise (i.e., split across columns), and hence it may need column-merging before downloading. The documentation suggests

```
SELECT ... FROM
    pdr3_wide.meas
    LEFT JOIN pdr3_wide.meas2 USING (object_id)
```

as the most effective way to do this.

Additionally, the sources should be filtered to only contain primary detections of objects, filtering out secondary detections in halos of tracts for instance – i.e., we would like to only have one row per astrophysical object, and duplicates should be removed with only the “best version” of the object in the database kept.

ID

The unique designation `object_id`, from `pdr3_wide.meas` and/or `pdr3_wide.meas2`.

Sky coordinates, uncertainty

The HSC `pdr3_wide.meas*` database does not contain a band-merged value for objects detected, instead giving best-fit positions for all filters separately. These are `*_sdsscentroid_ra` and `*_sdsscentroid_dec` (where `*_` is e.g. `g_`, or `i_` for all of the full `grizy` set of filters). Correspondingly there are `*_sdsscentroid_raerr` and `*_sdsscentroid_decerr` uncertainties for each of the five filters. These are all (along with `*_sdsscentroid_flag`) found in `pdr3_wide.meas2`. Additionally, to determine which filter’s measured position to use (see best detection, below) we will use `merge_peak_*` and `merge_footprint_*`, from `pdr3_wide.meas`.

To create a single astrometric uncertainty, in the absence of any covariance information, once the appropriate filter to take positions and uncertainties from is determined (see below), simply take the average of the appropriate filter values for `*_sdsscentroid_raerr` and `*_sdsscentroid_decerr`. Here we wish to take the geometric mean (instead of the arithmetic mean),

$$\left(\prod_{i=1}^{n} \sigma_i\right)^{\left(\frac{1}{n}\right)},$$

to preserve the ellipse area, so

```
sig_avg = sqrt(sdsscentroid_raerr * sdsscentroid_decerr).
```

Magnitudes, uncertainties

Again, we require magnitudes and corresponding uncertainties for all five filters. However, here we wish to use two separate algorithms for flux (and thus magnitude) determination, suited to different science cases: the `cmodel` magnitudes (the best option for non-crowded galaxy fluxes), from `pdr3_wide.meas`, and `psfflux` magnitudes (the best option for crowded stellar fields), from `pdr3_wide.meas2`.

In either case, we need to extract `*_Y_mag` and `*_Y_magerr`, where `*` is again any one of the filters (`grizy`) and `Y` is either `cmodel` or `psfflux`, giving five times two magnitudes and five times two magnitude uncertainties. We additionally need `*_Y_flag`, for a total of 3 columns per filter per algorithm.

We use `psfflux` within a Galactic latitude of 20 degrees ($|b| \leq 20$), and outside of the Galactic plane we use `cmodel`, for all five filters.

Best detection index

With HSC, we have independent positions for each source. Thus, unlike other datasets in which a combined position and uncertainty are reported, we need to decide which position to use. Hence we have two "best detection" index reduction tables for HSC: one to determine the best filter's position to keep, and another – the more general one, that we will propagate through to the reduced dataset – to determine the best magnitude to use within the cross-matching process.

Astrometry

The following quality cuts apply, with a default flag number of `2` unless criteria are met:

Astrometric Flag	Flag Number	Criteria
Non-detection	-1	`*_sdsscentroid_raerr` == NaN OR `*_sdsscentroid_decerr` == NaN OR `*_sdsscentroid_raerr` <= 0 OR `*_sdsscentroid_decerr` <= 0 OR `*_sdsscentroid_flag` is `True` OR `merge_footprint_*` is `False` OR `*_sdsscentroid_raerr` >= 1 OR `*_sdsscentroid_decerr` >= 1 OR `*_pixelflags_edge` is `True` OR `*_pixelflags_offimage` is `True`
Low quality (saturated)	0	`*_pixelflags_saturated` is `True`
Low quality (generic)	1	`*_sdsscentroid_raerr` >= 0.5 OR `*_sdsscentroid_decerr` >= 0.5 OR `*_pixelflags_bad` is `True` OR `*_pixelflags_clipped` is `True` OR `*_pixelflags_cr` is `True` OR `*_pixelflags_sensor_edge` is `True` OR `*_pixelflags_suspect` is `True`

For astrometry, priority in flags should be given to the one(s) that were been used to define the "footprint" – for each of the five filters, one or more of `merge_peak_*` should be True. If any of these filters has the default flag (`2` unless its quality was lowered by a quality-cut criterion), then it should be chosen – choosing in the order of `irzgy` (`i` first, `y` last). Otherwise, in same order of `irzgy` choose the positions and uncertainties defined from the first of the remaining `2` -flagged filters (i.e., the "best" filter with a flag of `2` and `merge_peak_*` being `False`). If you have not selected a filter, repeat this step (selecting the best `merge_peak_*` object then any objects with `merge_peak_*` of `False` , looping twice through in priority order) for `1` -flagged filters and then `0` -flagged filters. If somehow all filters are `-1` -flagged, the object should be totally discarded and no photometry selection made.

For reference, a (temporary) note should be made of which filter was selected astrometrically for the next step.

Photometry

The following quality cuts apply to the photometry, with a default flag number of `2` unless criteria are met:

Photometric Flag	Flag Number	Criteria
Non-detection	-1	`*_magerr` [1] == NaN OR `*_magerr` <= 0 OR `*_flag` is `True` OR `merge_footprint_*` is `False` OR `*_pixelflags_edge` is `True` OR `*_pixelflags_offimage` is `True`
Low quality (saturated)	0	`*_pixelflags_saturated` is `True`

Photometric Flag	Flag Number	Criteria
Low quality (generic)	1	`*_magerr` >= 0.5 OR `*_pixelflags_bad` is `True` OR `*_pixelflags_clipped` is `True` OR `*_pixelflags_cr` is `True` OR `*_pixelflags_sensor_edge` is `True` OR `*_pixelflags_suspect` is `True`

Any source with entirely negative flags is removed; if no filter is at least a zero-flagged detection then we assume the object should be discarded. For best detection filter, we prioritise first the filter the astrometry was chosen in, and then select filters in order of `irzgy`. Exactly as with the astrometry, first we check for `2` -flagged filters (astrometry filter, then the remaining four filters in `irzgy` order), then `1` -flagged filters and finally `0` -flagged photometry. It would be most sensible to define the order of the filters in increasing wavelength order, and hence have the columns in order of `grizy`, with `g` in `array[0]` (or `array[1]` for one-indexing) and `y` in `array[4]` (`array[5]`).

Note that we only record the photometric best detection index and not the astrometric-selection index, as the cross-match inputs require a single position-uncertainty record (and hence we filter the five positions down to one per source at this stage) but allow for multiple magnitudes per object.

Zone overlap index

A flag for "is this the primary zone this object is in, or is this object part of the padding around this zone", required for all cases where chunking is performed to parallelise the cross-matching process.

LSST Data Preview Columns, Issues

Given that the pipeline HSC uses to process its data is the LSST pipeline, we expect very similar columns and flags to be used, and have attempted to keep the two as closely joined as possible, such that the reduced tables created for HSC can serve as a test bed for LSST. However, there are a few outstanding issues with LSST Data Preview 0, and missing columns, that we briefly document here.

- DP0 does not provide `raerr` and `decerr` for its `*_centroid` algorithm, instead currently only giving `*_centroid.[x/y]` and `*_centroid_[x/y]err`. This should change as the pipeline matures and the propagation of pixel-level uncertainties to sky coordinates is implemented, but for now we do not have direct access to this information.
- `merge_footprint_*` is not included; `merge_peak_*` exists but apparently only for `merge_peak_sky` specifically. Is there a way within the LSST pipeline to know which the preferred filter was, or do we not need it for HSC and just fall back on `irzgy(u)` from the beginning?
- `*_cmodel` and `*_psfflux` do not directly quote magnitudes and uncertainties, only giving `_flux` and `_fluxerr` ; with zero-points these must be calculable but it would be nice if they were provided down the line.
- The more specific flags `[cmodel/sdsscentroid/psfflux]_flags_*` that HSC provides are missing (e.g. `*_sdsscentroid_flag_nosecondderivative` or `*_cmodel_flag_noshape`); we don't currently use them, but do we ever want to?
- Does LSST have the primary/secondary source issue, or are objects in a one-to-one correspondence between rows in tables and astrophysical sources?

[1] The use of `*_mag` , `*_magerr` , or `*_flag` in the photometric flag table refers to the selection of either `cmodel` or `psfflux` , and essentially `*_mag` , down from `*_Y_mag` (see magnitude discussion).