

1 Data Flow and Pipelines

- **Paranal pipeline:** data from VISTA are assessed for quality control (QC) in real time at the summit using a simplified data reduction pipeline. Because these reductions have to happen very rapidly and in a causal sequence, this pipeline relies on previously prepared library calibration information and hence can only be considered a first-pass result.
- **Garching pipeline:** the raw data are then collected onto USB discs, which are shipped to Garching. While there, the data are ingested into the ESO raw data archive and a second pipeline is run. This is used to monitor instrumental health, generate calibration information and to provide library calibration frames for the summit pipeline. Because time pressure is not so critical, more up-to-date calibration information can be applied and the quality control results are correspondingly better. The code for these ESO-run pipelines are written using the ESO CPL environment.
- **Cambridge pipeline:** once the archive ingestion is done, the USB discs are forwarded to Cambridge where the science data reduction is done. When running the science pipeline we are able to consider an entire night of data (or indeed a whole week of data) as a single entity and hence we can use information which is not available to the ESO pipelines. This leads to a much better result than can be obtained with the QC pipelines. The actual science pipeline code is written in C and Perl using the CASU pipeline infrastructure.

2 Data Processing and Monitoring

A great deal of work has to be done to near-infrared data before it is fully calibrated and free of instrumental signatures. Below is a brief description of the steps that are taken in the science processing pipeline.

- **Reset correction:** this is similar but not the same as a debias operation in CCD processing. Reset frames are taken for each exposure and are subtracted in the data acquisition system. Although this is not a pipeline reduction step per se, it is important to realise that this happens as it has an effect in estimating the linearity of the detectors.

- **Dark correction:** the dark current is estimated from a series of exposures taken with a dark filter inserted. Subtracting a mean dark frame also corrects several other additive electronic effects.
- **Linearity correction:** the VISTA detectors do not have a linear response. To estimate the non-linearity of each detector we need information on the readout timing, the exposure time and the reset image timing (this is because there is no shutter on the camera and in double-correlated sampling mode, the default, the reset frame is subtracted prior to writing images to disk).
- **Flat field correction:** dividing by a mean twilight flatfield image removes the small scale QE variations in the detector as well as the large scale vignetting profile of the camera. We also use the global flatfield properties of each detector to gain-normalise each detector to a common (median) system.
- **Sky background correction:** this removes the large scale spatial background emission that comes from the atmosphere as well as several remaining additive effects. The 2d background map is estimated using several different algorithms that combine the science images themselves with rejection or masking. Sometimes when large extended objects are present it is necessary to use 'offset sky' exposures to get a background map. Getting this right is one of the most difficult parts of near-infrared image processing.
- **Destripe:** the readout electronics for the VISTA detectors introduce a low-level horizontal stripe pattern into the background. Every exposure yields a different pattern, but groups of four detectors, readout through the same IRACE controllers, have the same pattern on a given exposure. This means there is a great deal of redundancy when it comes to estimating the stripe pattern.
- **Jitter stacking:** infrared detectors often have large numbers of cosmetic defects implying infrared imaging is invariably done in a jitter mode, whereby an observation of a region is broken up into several shorter exposures and the telescope moved slightly between them. At this point in the reduction the jitter series is shifted and combined to form a single image stack, using positions of detected objects on all the

detectors to compute the shifts. This allows bad pixel regions in one exposure to be rejected in favour of good pixels in other exposures.

- **Catalogue generation:** information on astronomical objects on the stacked image are extracted at this point. The parameters include positions, fluxes, and shape descriptors which are combined to generate object morphological classifications and form the basis for calibration and QC information.
- **Astrometric and photometric calibration:** objects on the catalogues are matched by their counterparts in the 2MASS point source catalogue. Because 2MASS has such a high degree of internal consistency it is possible to calibrate the world coordinate system of VISTA images to better than 50 milli-arcseconds. The 2MASS magnitudes are also used in conjunction with colour equations to provide photometric zeropoints to an accuracy of 1-2% (depending upon the wave band).
- **Tiling:** the overlap areas in VISTA pawprints needed to form a contiguous tile are large, so to achieve full depth in most observations, the pawprints must be combined into tiles. Once this is done the catalogue step is repeated.

It is worth noting that during commissioning and testing of the instrument we looked for evidence of detector crosstalk and sky fringing and could find none. Hence it has not been necessary to deal with them in the pipeline. We also looked for evidence of persistence, which manifests itself as a glow on a detector where a bright object was recently observed. For the VISTA detectors this turned out to be a very small effect which only occurred when extremely bright stars (which are rare) are observed. In practice this effect is negligible and is ignored during pipeline processing.

The processing of data is monitored using web pages that are automatically updated when new raw data are received or when a night of data is processed (see <http://casu.ast.cam.ac.uk/surveys-projects/vista/data-processing/>). The catalogue information forms the basis for the various QC plots and summary statistics.