

Scanamorphos user guide

H. Roussel

Institut d'Astrophysique de Paris, Université Pierre et Marie Curie (UPMC), CNRS (UMR 7095), 75014 Paris, France
e-mail: roussel@iap.fr

software webpage: <http://www2.iap.fr/users/roussel/herschel/>

Scanamorphos is public software available to post-process scan observations performed with the Herschel photometer arrays. This post-processing mainly consists in subtracting the total low-frequency noise (both its thermal and non-thermal components), masking cosmic ray hits, and projecting the data onto a map. Although it was developed for Herschel, it is also applicable with minimal adjustment to scan observations made with other bolometer arrays, provided they entail sufficient redundancy; it was successfully applied to P-Artémis, an instrument operating on the APEX telescope. Contrary to most other algorithms among those aiming to subtract the low-frequency noise on most timescales, i.e. maximum-likelihood softwares and high-pass filters, *Scanamorphos* does not assume any particular noise model, and does not apply any Fourier-space filtering to the data, but is an empirical tool using purely the redundancy built in the observations – taking advantage of the fact that each portion of the sky is sampled at multiple times by multiple bolometers. It is an interactive software in the sense that the user is allowed to optionally visualize and control results at each intermediate step, but the processing is fully automated.

This document serves as a brief user guide to *Scanamorphos*. A description of the algorithm can be found in the companion paper. Additional usage advice, for example for specific types of observations, can be found on the software webpage. The full list of options with their definition is included in the README file of the distribution and recalled here on page 4. Practical examples following those of the companion paper are given in a subdirectory.

1. Input data and parameters

Input data are in the form of structures, one for each scan. For SPIRE, data are recorded for all three arrays simultaneously, so the structures contain the signal and coordinate series of all arrays, and these share the same time vector. For PACS, only two filters can be used simultaneously (one on each array), and in addition the processing up to level 1 may not produce the same time vector for them; consequently, the data of each filter and each scan are stored in a separate structure. Utilities are provided to convert HIPE output files (in fits format) to *Scanamorphos* input data. The SPIRE pipeline works scan leg per scan leg and usually outputs one fits file per scan leg, instead of one file per scan. This is accounted for, and one must only make sure that all the HIPE files corresponding to a single scan are given names sharing a unique root. Example HIPE scripts to reduce SPIRE data up to level 1 and save HIPE files with the correct format are provided with the *Scanamorphos* distribution. The PACS pipeline produces fits files that can contain several scans taken in the same direction (several repetitions). We also provide a utility to convert these files to input structures, and slice the data into separate scans.

The list of input file names has to be entered in an ascii file. Up to version 23, care must be taken that in this list, successive scans are in alternating directions, to ensure that the destriping algorithm works as expected. When scans are considered in chronological order, this is the default sequence for SPIRE nominal observations, but in general not for PACS observations. An example is given explicitly in the software distribution for all versions up to 23. From version 24 on, the ordering of the scans on input is no longer relevant, since the code figures out how they have to be rearranged. The imprint and direction of the scans can be visually checked with the `/vis_traject` option.

The choice of the pixel size and orientation of the final maps is left to the user. The pixel size can be different from the default if, for example, it is wished to produce congruent maps at different wavelengths. However, not all choices will result in adequate coverage of the field of view or good sampling of the point spread function. For parallel-mode observations using the large scan speed (60 arcsec/s), or parallel-mode observations using the blue array of PACS, that have much lower coverage and redundancy than nominal-mode observations, it is advisable to use pixels sizes of FWHM/2 instead of FWHM/4.

It may also be desirable to build maps in scan coordinates rather than right ascension and declination, for very elongated fields of view for instance, or when one wishes to use point response functions oriented in the same way in combination with the target maps; this option is implemented as the default.

The user has control over the inclusion of turnaround data (data at field of view edges, acquired while the telescope acceleration is non-zero). Turnaround data are always used in the drifts determination, whenever they are included in the input data, but they are weighted differently, so that they do not have artificially high coverage values; they can however be excluded before the final projection.

2. Main options

A number of command-line keywords and parameters are available to control the degree of interaction, and more importantly the processing:

- `/jumps_pacs`: If PACS data are affected by brightness discontinuities (see Section 3.6.1 of the companion paper), this option can be used to detect the discontinuities and mask the affected samples. When *Scanamorphos* is run in interactive mode, the user is required to visually check, and confirm or infirm each detection. The very unstable rows or bolometers are automatically detected and pre-filtered by the same module.
- `/nothermal`: If the thermal drift subtraction performed within HIPE (for SPIRE) is believed to be correct – or the small-scale thermal drift is believed to be negligible and it is wished to reduce the processing time for a very large field – the corresponding step in *Scanamorphos* can be bypassed.
- `/noglitch`: Likewise, if the deglitching performed within HIPE is trusted, it is possible to not perform it in *Scanamorphos*.
- `/nogains`: For SPIRE, if the HIPE version used locally does not match the flux calibration version for which the relative gains were derived, or if the gain correction was already performed within HIPE, the gain correction must not be performed in *Scanamorphos*. For PACS, if the distortion flatfield (calibrating the factors between physical detector sizes and fiducial detector sizes) was already applied within HIPE with the `convertToFixedPixelSize` task, then it must not be applied a second time in *Scanamorphos*.
- `/parallel`: For observations acquired in parallel mode, it is necessary to select the relevant option, because the smaller sampling rate implies that the frequency intervals used to compute the high-frequency noise amplitudes have to be adapted.
- `/galactic`: For observations of bright Galactic fields, it is highly recommended to use the relevant option, so that no sky structure can be removed by the baselines, as explained in Section 3.4 of the companion paper. It should *not* be used for observations of diffuse Galactic structures, and in general for fields where the brightness gradients induced by the low-frequency noise are completely dominant over genuine brightness gradients of the sky. In case of doubt, it is advised to produce maps both with and without the `/galactic` option, and to compare them to decide which is the best strategy.
- `/minimap`: For fields observed in the mini-scan or small-map mode, that produces very small maps by definition, the relevant option should be selected in order to deactivate the destriping and average drift subtraction modules, since there are not enough resolution elements with nominal coverage across the map to make these corrections necessary and accurate. These corrections may also need to be deactivated (by selecting the `/minimap` option) when there are only two or three scan legs per scan. As above, in case of doubt, it is advised to produce maps both with and without the `/minimap` option and to compare them.
- `/flat`: This option forces the sky background to be flat, when the observation strategy does not allow a correct derivation of baselines. The case that motivated the introduction of this option is that of an extended source greater than the field of view, and observed with a set of scan position angles concentrated around its minor axis, severely limiting the area covered with enough redundancy to robustly estimate long-timescale drifts.
- `/nocross`: For observations in which the nominal coverage consists of scans taken in only one direction, some adaptations are necessary, and in particular no destriping can be performed. Note that such observations are not recommended. This option is provided to qualitatively assess observations that are not completed yet, but is not guaranteed to produce science-grade maps.

In visualization mode (with the `/visu` or `/debug` option), it is possible to inspect intermediate results, in particular the average drift series, and for the latter choose to apply the correction or not. When the nonlinear component of the thermal drift is very small (with respect to the high-frequency noise), the correction is not necessary. From version 24 on, it should no longer introduce low-level noise, since only the parts of the drift with a high significance level are retained.

3. Memory usage control

For very large data volumes, two parameters controlling memory usage can be modified directly inside the main program (but only experienced users should). One controls the maximum size of the drift difference matrix; if this threshold is exceeded, then the minimum timescale of the average drift is increased (the minimum timescale of the uncorrelated drifts is left intact, since it has a much smaller impact on memory usage). The second controls whether the field of view has to be sliced into several overlapping spatial blocks, that are processed separately and stitched together at the end of the processing; this is useful when the data volume is too large to be treated as a single block, and when the scientific interest in not in very extended sources (e.g. for deep surveys of high-redshift galaxies). To prevent slicing the field into several blocks, the `max_n_samples` parameter has to be increased (see the online “tips page”).

A command-line option explicitly allows the number of spatial blocks to be specified, starting with version 5. This option was introduced to allow the time resolution for the average drift computation to be increased (see Section 3.8 of the companion paper), but should not be used if the field of view contains bright and very extended emission (i.e. extended over more than a few percent of the total field size).

4. Non-default astrometry

By default, the astrometry and spatial grid are determined from the input scans. The user has however the latitude to change the spatial grid or request that only a portion of the entire field be processed and mapped:

- 1) by supplying a reference fits header (in the form of an IDL string array). In that case, the astrometry is taken entirely from the header. With this option, only one array should be processed at a time. Note that it is now usable in batch mode, and then requires the header to be in the form of an IDL save file, containing a variable named exactly `hdr_ref`.

2) by supplying a three-element array containing the central coordinates and minimum radius of a subfield to be excised from the data (all in degrees):

```
cutout=[ra_center, dec_center, radius]
```

In case 1, if the reference image from which the header has to be extracted has been produced by *Scanamorphos* (as the first plane of the fits cube), and if it has to be rotated first by a given angle, here is the sequence of IDL commands to create the reference header `hdr_ref`:

```
cube = readfits('ref_ima.fits', hdr_cube)
hdr_ima = hdr_cube
sxaddpar, hdr_ima, 'NAXIS', 2
sxdelpar, hdr_ima, 'NAXIS3'
hrot, cube(*,*,0), hdr_ima, ima_rot, hdr_ref, angle, -1, -1, 1
```

where *angle* is the value of the rotation angle in degrees, clockwise. To save the reference header to disk, this command can be used:

```
save, filename='hdr_ref.xdr', hdr_ref, /xdr
```

It is also possible to specify astrometric offsets for each scan (angular distances in arcsec) with the `offset_ra_as` and `offset_dec_as` parameters.

5. Output

As the successive processing steps are run, some information is printed to the terminal window (and to a log file if batch mode is used). In particular, before the final projection, the processing time and drift amplitudes are indicated.

For those who are familiar with IDL, regular stops in the processing occurring in visualization mode enable the user to interrupt the code and temporarily switch to the command line, in order to check the content of variables or re-display images and graphics with different dynamic ranges. Altering variables is not recommended, unless a deep knowledge of the data and algorithm have been gained beforehand.

On output, the signal, error, drifts and weight maps are assembled into a cube, of which the third dimension is the plane index. This cube and the associated astrometry are saved on disk in a file conforming to the Flexible Image Transport System (FITS) standard (Wells et al. 1981; Hanisch et al. 2001).

Some cubes have a fifth plane, in that case containing a “clean” signal map built with each scan weighted by its inverse variance in the corresponding pixel. It is provided to allow easy identification of residual artefacts (by comparison with the first plane), but in principle it should not be used for photometry.

The weight map should always be inspected, since it allows to locate pixels affected by saturation or abusive deglitching (that have significantly lower weight values than their neighbors) and to select the part of the signal map with nominal coverage.

6. To combine overlapping fields

Overlapping fields can be combined either during the processing or after the final maps have been created.

If all the fields have been observed with the same settings (same scan speed and sampling rate), it is usually best to process them together. If the area of overlap is a large fraction of the total area, then the processing is standard.

In the opposite case, when the fields are largely disjoint, this is automatically detected during the destriping, and the brightnesses of all scans are matched in their area(s) of overlap. Note that this matching may not be perfect if the `/galactic` option is used, and in any case it is impossible to recover the general gradient across distinct fields. When the combined field is very large, it is advised to use the `/nothermal` option to keep the processing time within reasonable bounds. It is also strongly recommended to use the `nzadata='no'` option, so that noisy data on scan edges do not contribute to central parts of the map.

For observations taken at different scan speeds or sampling rates, it is mandatory to process them separately. The final maps can then be combined with the `stitch_blocks.pro` routine, provided they have been created with the same reference header and have been renamed appropriately. Care must also be taken that they are arranged in the optimal order, such that the overlap between successive maps is maximal. For example, if we wish to use the root name `field` for `n = 3` maps, we will rename them:

```
field_0.fits, field_1.fits, field_2.fits
```

and they are combined with these commands:

```
hdr_ref = headfits('field_0.fits')
stitch_blocks, nblocks=n, root_out='field', hdr=hdr_ref
```

which will produce a file named

```
field_combined.fits .
```

References

Hanisch, R.J., Farris, A., Greisen, E.W., et al. 2001, *A&A*, 376, 359
Wells, D.C., Greisen, E.W., & Harten, R.H. 1981, *A&AS*, 44, 363

Table 1. Summary of inputs and options

auxiliary files	
scanlist_spire	ascii file containing the directory and file names of the input scans
scanlist_pacs	idem
command line keywords and parameters	
spire	select to process SPIRE observations
pacs	select to process PACS observations
visu	select to visualize intermediate results and stop after each major step
vis_traject	select to visualize the imprint and direction of all scans
debug	detailed visualizations, for experts only
version	select to print the software and SPIRE relative gains versions
jumps_pacs	select to detect and mask brightness discontinuities in PACS data ^a
nothermal	select to bypass the short-timescale average drift correction
no glitch	select to bypass glitch masking and asteroid detection
no gains	SPIRE: select if the gain version does not match the local flux calibration version or if the relative gain correction has already been applied in HIPE PACS: select if the distortion flatfield has already been applied in HIPE
paralle	select in case of parallel-mode observations
galactic	select in case of very bright Galactic field or similar dataset
minimap	select in case of mini-scan or small-map mode
flat	select to force the sky background to be flat
nocross	select for observations consisting of scans in a single direction or non-overlapping
hdr_ref	reference header to enforce the same astrometry ^b (IDL string array or IDL save file containing a variable named “hdr_ref”)
cutout	central RA and DEC coordinates and radius of the subfield to be processed and mapped
nblocks	number of spatial blocks the field of view will be sliced into ^c
block_start	index of the first spatial block to be processed, when nblocks > 1 ^c
one_plane_fits	select to save each plane in a separate fits file
offset_ra_as	right ascension offsets (angular distances) to be applied to the map coordinates (one for each scan)
offset_dec_as	declination offsets (angular distances) to be applied to the map coordinates (one for each scan)
inputs at the prompt	
output directory	directory where final maps and intermediate variables will be stored ^d
output FITS file root	full file name = file root + instrument + wavelength + block index + “.fits”
detector arrays	choose only one for PACS and between one and three for SPIRE
map orientation	binary choice: orientation of the first scan (default), or standard astronomical orientation
pixel size	in arcsec, equal to FWHM/4 by default
turnaround data	option to reject turnaround data from final map

Notes. ^(a) In interactive mode (with /visu), the user is allowed to reject any false detections. ^(b) The field of view of the reference header must of course be the same as that of the input scans. ^(c) The option to slice the field of view should be used only in very specific cases: see Section 3.8 of the companion paper. When the first blocks have already been processed, if the program is interrupted, it is possible to restart from the first unprocessed block with the `block_start` parameter. ^(d) Make sure to have enough space in this directory to store all the intermediate variables necessary for the processing, and to be able to write files in it.