100,000 Molecular Clouds Across the Main Sequence: GMCs as the Drivers of Galaxy Evolution

A *Physics at High Angular resolution in Nearby GalaxieS* (PHANGS) Survey
Data Delivery Document for PHANGS-ALMA

ALMA Large Program PI: Eva Schinnerer (MPIA/Heidelberg)
ALMA Large Program ID: 2017.1.00886.L
ALMA Pilot Projects: 2012.1.00650.S (PI: Schinnerer), 2015.1.00925.S (PI: Blanc), 2015.1.00956.S (PI: Leroy), 2018.1.01651.S (PI: Leroy)

Reference Publications:

Survey Paper: Leroy et al., (2021a), ApJS, in press (ADS)
Pipeline Paper: Leroy et al., (2021b), ApJS, in press (ADS)

Table of Contents

1. Introduction

This is the first public delivery of data for 74 nearby galaxies within 23 Mpc where the PHANGS-ALMA large program and individual pilot projects (see list below) mapped $^{12}CO(2-1)$ emission over a large area of their star-forming disks. For all targets, the data consist of combined 12-m, 7-m, and total power array mapping with typically 1.3” angular resolution and 2.5 km/s velocity resolution. The median area mapped for a typical target is 7 arcmin^2, covering on average 70% of the recent star formation visible from IR and UV maps. Details on the PHANGS sample and ALMA processing can be found in Leroy et al. (2021a,b).

This release is tagged v4.0 because it follows earlier internal deliveries to the PHANGS team. Incremental updates to this release that make minor improvements and use the same basic processing will be labelled v4.1, v4.2, etc. If the PHANGS team implements significant processing improvements in the future, the associated data products will be released using higher base version numbers (e.g., v5.X).

The PHANGS ALMA v4.0 release includes data from three pilot programs and the PHANGS-ALMA Large Program itself. The PHANGS team is also releasing data from several follow-up programs archival data sets that meet the target selection and observational quality criteria for the PHANGS-ALMA Large Program. These latter data can be obtained at the team data page[1]. The versioning of this release will remain consistent, i.e., v4.0 includes the archival and follow-up observations.

This README is part of the PHANGS-ALMA delivery to JAO. The data delivered to JAO and released via the ALMA archive must comply with observatory policy. The JAO release thus includes ONLY the data from the pilot programs and the PHANGS-ALMA Large Program itself. Similar releases are also hosted on the VOSpace file system operated by CANFAR[2] and through the Canadian Astronomy Data Centre (CADC) archives. The content of the files in the JAO release is identical to the other archive deliveries, though some metadata may be different to conform to specific archive guidelines. Data released via CADC will contain some targets from archival programs that are imaged using the PHANGS imaging pipeline but are not formally part of the Large Program.

2. Content of the Data Delivery

We deliver products for 12m+7m+TP feathered data in FITS format. The products can be read using the standard FITS readers (e.g., the `astropy` package[3])

The data delivery contains position-position-velocity cubes, signal identification masks, local estimates of the noise, and two-dimensional maps that are generated from the data cubes cubes using two different signal identification schemes. Maps of the uncertainties associated with the two-dimensional emission maps are also distributed. The data cubes and derived two-dimensional maps are delivered at the native resolution for each cube.

2.1 Data Product Types

Data cubes (`cubes`)

We provide the cubes at their native round-beam resolution in equatorial ICRS coordinates and the LSRK velocity reference frame. The cubes are primary beam corrected and converted to main beam brightness temperature units (Kelvin).

[1] http://phangs.org/data
[2] https://canfar.net/
[3] https://docs.astropy.org/en/stable/io/fits/

Strict mask products (`strict_mom0`, `strict_mom1`, etc.)
This directory contains two-dimensional maps that have been generated from data cubes after applying a "strict" signal identification mask. This "strict" mask only includes emission that is identified as signal with high confidence in the data cubes. However, because of the stringent signal identification criteria, these maps typically include less of the total flux (as determined from the total power data), especially at higher resolution. Details of the masking strategy and completeness statistics are presented in the PHANGS pipeline paper (Leroy et al., 2021b). Use of the strict maps is advised when low noise is required, some incompleteness can be tolerated, and a very reproducible calculation is desired.

Broad mask products (`broad_mom0`, `broad_tpeak`, etc.)
This directory contains two-dimensional maps that have been generated from data cubes after applying a ``broad'' signal identification mask. These broad masks include all sightlines where signal is identified at any resolution. The broad masks have excellent completeness and larger covering fractions than the strict maps. However, because these masks include more regions with faint emission or regions near bright emission they appear noisier and can contain false positives. Details of the masking strategy and completeness statistics are presented in the PHANGS pipeline paper (Leroy et al., 2021b). Use of the broad maps is advised when high completeness is required, and some additional noise can be tolerated. Because higher order moment calculations (e.g. moment-2) become unstable in the presence of noise, we do not deliver broad mask versions of these map products. The mom1wprior velocity field map, which is constructed using information from both the strict and broad masks, is bundled with the broad mask products.

2.2 Filename convention

The filenames of PHANGS-ALMA data products follow a common convention : *group.archiveprefix*.*LPidentifier*.*galaxy*_*arraycombination*_*linetype*[_*masktype*][_*maptype*].*fits*
Example:
group.uid___A001_X2fe_X30b.lp_schinner.ngc1433_12m7mtp_co21_broad_emom0.fits

- *Archiveprefix* -- Each file has an ALMA archive specific prefix set by its Group Observing Unit Set (GOUS). This indicates the archive data product from which these data are derived. Note that some targets are the mosaics of multiple GOUSs. We deliver identical data for each of these GOUS.
- *LPidentifer* -- The large program identifier `lp_schinner` is the same for all targets.
- *galaxy* -- we use NGC numbers for the names of PHANGS ALMA targets whenever possible. The remaining galaxies are identified using their IC number.
- *arraycombination* -- for this release, we deliver data products constructed from the 12m+7m interferometric data feathered with the total power data. We indicate these as "`12m7mtp`" respectively in the filenames. This distinguishes the files from data produced with only the 7m interferometric array, which are included in other releases.
- *linetype* -- for this release, we deliver $^{12}CO(2-1)$ data only, which we indicate as "co21" in the filenames.

- *masktype* -- strict or broad, to distinguish maps generated using the two signal identification procedures described above. The masktype is not specified in the filename of the mom1wprior version of the velocity field maps, as they are constructed using information from both the strict and broad masks.
- *maptype* -- The different two-dimensional map types generated from the PHANGS-ALMA CO(2-1) cubes are described below. Error maps are indicated by adding an 'e' before the maptype, e.g. '`emom0`' indicates the error associated with the moment-0 ('`mom0`') map.

2.3 Data Products

Please see Leroy et al. (2021a,b) for full details on data product generation.

- Emission cubes
    - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21.fits`
    - Unit: K
    - Short description: CO(2-1) spectral line cubes

- Noise cubes
    - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_noise.fits`
    - Unit: K
    - Filename map type: noise
    - Short description: position-position-velocity cube containing the position- and frequency-dependent estimate for the noise in the CO(2-1) spectral line cube. The noise estimation procedure is described in the PHANGS pipeline paper (Leroy et al., 2021b).

- Signal identification masks (cubes)
    - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_broadmask.fits`
    - Unit: binary mask
    - Filename map type: broadmask or strictmask
    - Short description: position-position-velocity cube with same dimensions as the corresponding CO(2-1) data cube. 1 = pixel with significant emission, 0 = no significant emission. Strict masks are determined by finding all emission with a brightness $>4\sigma$ in two adjacent channels and expanding this initial selection to all connected regions with emission $>2\sigma$ in two adjacent channels. Here $\sigma(x,y,v)$ is the noise cube specified above.

- Integrated intensity (moment-0) and associated uncertainty

- Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_broad_mom0.fits`
- Unit: K*km/s.
- Filename map type: mom0
- Short description: This is the direct integration of the cube along the velocity axis inside the relevant mask. The uncertainty comes from error propagation assuming independent velocity channels and using the empirical noise estimates.

- Peak brightness temperature, peak brightness temperature in 12.5km/s window and associated uncertainties
  - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_broad_tpeak.fits`
  - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_broad_tpeak12p5.fits`
  - Unit: K
  - Filename map type: tpeak, tpeak12p5
  - Short description: Maximum intensity along the velocity axis for each line of sight. For tpeak12p5, this is the maximum intensity along the velocity axis for each line of sight after smoothing the cube with a 5 channel boxcar along the spectral axis (but not downsampling). Note that for the tpeak and tpeak12p5 calculations, the mask used covers all spatial pixels. In each pixel, it takes the peak intensity considering all channels where the signal mask has any coverage.

- Intensity weighted mean velocity (moment-1) and associated uncertainty
  - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_strict_mom1.fits`
  - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_mom1wprior.fits`
  - Unit: km/s
  - Filename map type: mom1, mom1wprior
  - Short description: The strict_mom1 maps are the intensity-weighted mean velocity calculated using pixels inside the strict signal-identification mask. The mom1wprior maps additionally include moment-1 values calculated using pixels within the broad mask that satisfy three criteria: (i) there is no strict mask measurement for that pixel, (ii) the integrated intensity value at that pixel is detected with a signal to noise greater than 2, and (iii) the measured velocity is within 30 km/s of the intensity-weighted mean velocity estimated from the PHANGS-ALMA CO(2-1) data at 15” resolution. For sightlines with a single velocity component, the moment-1 reflects the average velocity of the emission.

In the case of line profiles with two or more velocity components, the moment-1 will sit intermediate between the components.

- Line width metrics and associated uncertainties.
  - Velocity dispersion / "moment-2"
    - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_strict_mom2.fits`
    - Unit: km/s
    - Filename map type: mom2
    - Short description: This is the intensity-weighted second moment, measuring the rms scatter about the intensity weighted mean velocity. This metric is highly sensitive to the inclusion of noise, so only a strict mask version of this map is delivered. For sightlines with multiple velocity components, it will also be sensitive to the velocity separation between peaks. In that sense, convergence between this and the equivalent width (see below) is a crude diagnostic of Gaussianity.

  - Equivalent width
    - Example filename: `group.uid___A002_X5a9a13_X57a.lp_schinner.ngc0628_12m7mtp_co21_strict_ew.fits`
    - Unit: km/s
    - Filename map type: ew
    - Short description: This is the integrated intensity (moment-0) divided by the peak intensity. That is, this is the rectangular width needed to supply the full line width at the peak intensity. A prefactor is then applied to recast as the equivalent sigma for the case of a Gaussian line profile. This is a highly robust statistic in the sense that it behaves well in the presence of noise or multiple components. It can easily miss subtleties in the line profile and has some dependence on spectral resolution.

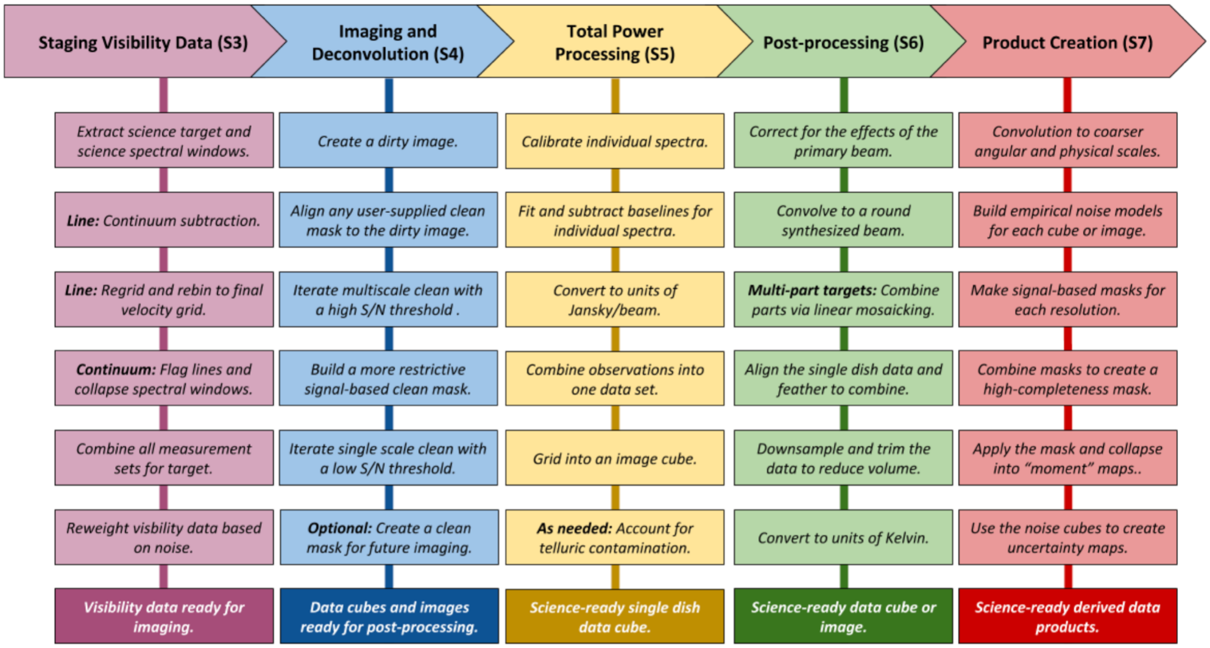
Caveat: No line width maps provided are yet corrected for the line spread function (channel width + channel-to-channel correlation) nor are they corrected for biases due to finite sensitivity (this is a particular issue for moment-2).

3. Methods

The full set of data can be reproduced from calibrated measurement sets using the PHANGS-ALMA imaging and analysis pipeline. The pipeline is described in Leroy et al. (2021b) and is available on github[4]. The pipeline configurations required to completely reproduce the PHANGS-ALMA release are also available in that repository. The figure below, adapted from

[4] https://github.com/akleroy/phangs_imaging_scripts

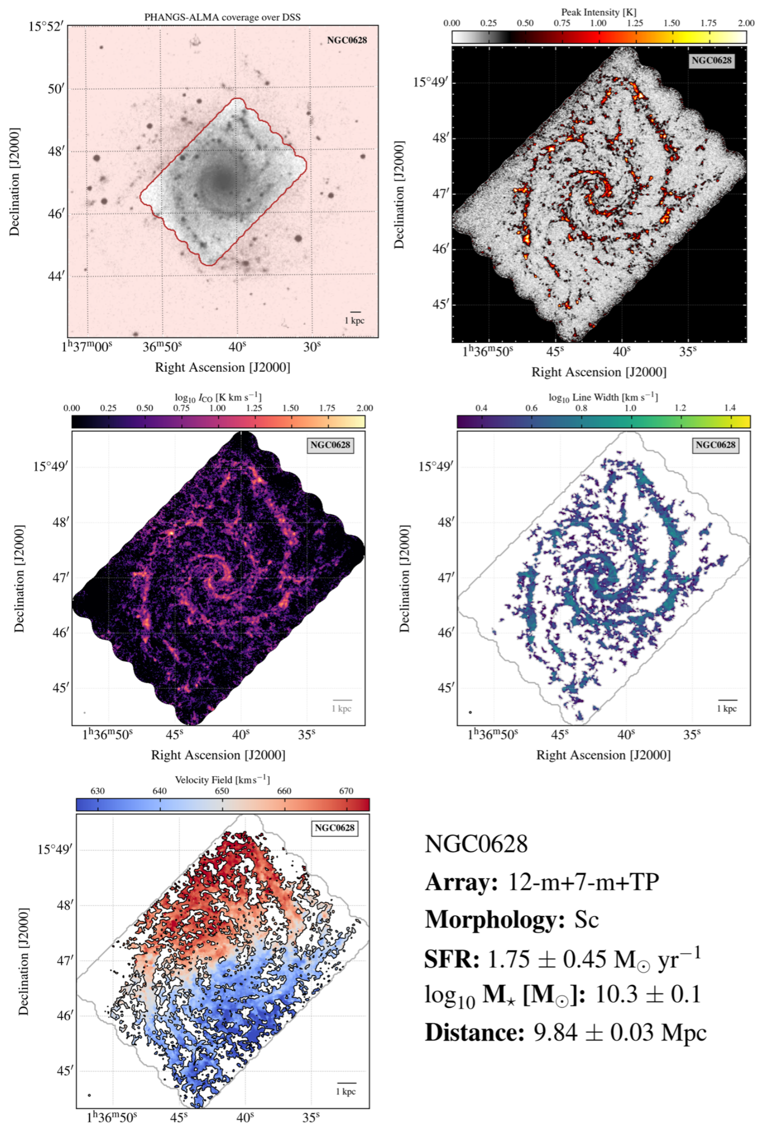
Leroy et al. (2021b), illustrates the action of the pipeline.



The numbers in the headings indicate the relevant sections of that paper which lay out the algorithm in each stage of the pipeline in detail.

4. First look

A full atlas of the PHANGS-ALMA data set is available at the PHANGS data page[5]. The figure below shows an example galaxy from the data set. The figure shows (left-to-right, top-to-bottom) (a) the Digital Sky Survey image of the target galaxy with the observed region highlighted, (b) the peak intensity map of the galaxy, (c) the $\log_{10}$ of the integrated intensity map of the galaxy using a broad mask, (d) the $\log_{10}$ of the line width estimated through the equivalent width method with a broad mask, and (e) the line-of-sight velocity estimate using the broad mask and a low-resolution prior velocity field. The data in the lower-right are the values that characterize the galaxy adopted by the PHANGS survey team for this data release.



[5] http://phangs.org/data

5. Quality Assurance

Section 8 of Leroy et al. (2021b) lays out the quality assurance process used by the PHANGS-ALMA team. Quality Assurance occurred through three channels summarized below:

- *Manual Inspection*: several members of the PHANGS team engaged in manual data inspection, reviewing all data products and statistical summaries of image properties to identify low-quality images or errors in pipeline processing. All issues were documented on a galaxy-by-galaxy or product-by-product basis. There were repeated rounds of manual inspection and revision until all issues were retired.
- *Regression Tests*: The collaboration used versioned data sets and created a regression framework for comparing the results of the imaging pipeline to previous versions of imaging. These regression tests checked both variations in the data and the metadata, assessing their integrity and identifying cases of large variation from version to version.
- *End-to-End Pipeline Tests*: The team also created test data sets using the ALMA simulator in CASA to generate images similar to the actual PHANGS ALMA data with a known ground truth. These images are processed end-to-end through the pipeline leading to images with similar characteristics as the actual survey data. These tests are used to assure that the pipeline is functioning as intended and to assess the impact of our choices in the imaging and deconvolution process.

6. Known Issues

We strongly encourage users to consult Leroy et al. (2021a,b) where many of the known issues below have been identified and quantified. If you identify more issues, please send a message to almahelp[at]phangs.groups.io

6.1 Generic issues

- As discussed above, we do not separately distribute 12-m only, 12-m+7-m only, 7-m only, or TP-only data sets. This release includes only our best single image for each galaxy. These individual-configuration images may be made available based on request.
- Many galaxies have been observed in two or more parts and then linearly mosaicked after imaging. In these cases, the different parts may have different surface brightness sensitivity, particularly when the two parts were observed in different configurations. Further, the region of overlap between the two maps tends to be more sensitive as a result of being observed in both data sets. The delivered data have a single matched synthesized beam, but users should be aware of the sensitivity variations.
- Faint continuous negative features at a level lower than the rms noise can still be found around some bright sources. In general, these “bowling” features are mitigated compared to most other imaging approaches and further reduced by feathering with total power observations. Nonetheless some sources, especially bright galaxies imaged with only the 7-m+TP arrays, are affected.
- In some cases the galaxies show noise gradients across the map and inhomogeneous primary beam coverage. In all cases that we examined, these variations indeed mapped back to uneven observational coverage, i.e., disparate time spent in different parts of the map.

- Most cubes show a noise gradient as a function of velocity, with total magnitude 20-30% variation across the cube. The effect is described in detail in the data processing and pipeline paper (Leroy et al., 2021b).
- The first and last channel of some of the $^{12}CO(2\text{-}1)$ cubes are filled with not-a-number (NaN) values. This occurs naturally during the processing and does not reflect a problem with the data, but users may wish to be aware of this during their analysis.

6.2 Galaxy-specific issues

- A few galaxies are known to have cubes with weak signal, which can also lead to apparently empty moment maps. This reflects the real faintness of CO emission, not a problem with the data. Galaxies known to have faint CO emission include: IC 5332, NGC 1809, NGC 2283, NGC 3239, NGC 4496A, NGC 4571, NGC 5042, and NGC 5068. Flux recovery in the broad moment maps compared to the total power data varies across this list from an almost-empty moment map to faint, noisy, but detected signal.
- NGC 6744 has an unusual coverage pattern, with two separated regions to the north and south of the bulge of the galaxy. The middle strip has not been observed.

7. Acknowledgements

If you use the PHANGS ALMA v4.0 data delivered via JAO please include the acknowledgments, project codes, and references below.

Please first cite both the PHANGS-ALMA survey paper: A. K. Leroy et al. (ADS) ApJS, in press and the PHANGS-ALMA data pipeline paper: A. K. Leroy et al., (ADS), ApJS, in press. PHANGS-ALMA was only possible through the dedicated effort of several people over half a decade. The appendices of these two papers outline the contributions of the individual team members.

Please also add the following acknowledgments

- This paper makes use of the following ALMA data:
  ADS/JAO.ALMA#2012.1.00650.S,
  ADS/JAO.ALMA#2015.1.00925.S,
  ADS/JAO.ALMA#2015.1.00956.S,
  ADS/JAO.ALMA#2017.1.00886.L,
  ADS/JAO.ALMA#2018.1.01651.S

- ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

- If you are using only a subset of the data, you can optionally use the following chart, adapted from Table 2 of the PHANGS-ALMA survey paper to select which programs to acknowledge:

  ADS/JAO.ALMA\#2012.1.00650.S : PI Schinnerer (pilot), NGC 0628 (M74)
  ADS/JAO.ALMA\#2015.1.00925.S : PI Blanc (pilot), 9 galaxies
  ADS/JAO.ALMA\#2015.1.00956.S : PI Leroy (pilot), 8 galaxies
  ADS/JAO.ALMA\#2017.1.00886.L : PI Schinnerer (large program), 54 galaxies,

8. Help and Feedback

A comprehensive description of the PHANGS-ALMA sample is available in the survey paper (Leroy et al. 2021, ApJS, accepted, ADS) and a comprehensive description of the processing (calibration, imaging, and product creation) is available in the pipeline paper (Leroy et al., 2021, ApJS, accepted, ADS).

We welcome feedback and suggestions on data products and documentation. Please direct suggestions to the PHANGS ALMA data reduction group (almahelp[at]phangs.groups.io).