PHANGS-ALMA Data Delivery README
2021-04-15

1. PREAMBLE

This is the first public delivery of data for 90 nearby galaxies within 23 Mpc where the PHANGS-ALMA large program and individual archival projects (see list below) mapped 12CO(2-1) emission over a large area. For most 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. For a subset of very nearby targets, the survey used only 7-m and TP mapping but still achieved <~ 150 pc resolution, matched to the rest of the survey. Details on the PHANGS sample and ALMA processing can be found in Leroy et al. (2021a,b).

This initial data release will also be distributed via the ALMA Archive [1], the JAO Large Program page[2] and the PHANGS collection at CADC[3] (n.b., these links will be updated as releases are completed).

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

[1] https://almascience.nrao.edu/asax/
[2] https://almascience.nrao.edu/alma-data/lp
[3] https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/

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 internal release includes data from 1) several pilot programs, 2) the PHANGS-ALMA Large Program itself, 3) several follow-up programs, and, 4) several archival data sets that meet the target selection and observational quality criteria for the PHANGS-ALMA Large Program. The versioning of this release will remain consistent, i.e., v4.0 includes the archival and follow-up observations.

Some data products in a PHANGS-ALMA data release are associated with a specific version of the PHANGS sample table, which describes the main characteristics of the target galaxies including distance. For this release, the associated sample table is: phangs_sample_table_v1p6.fits. This file is copied into the release.

2. PHANGS vs CADC vs JAO releases

This README is part of the PHANGS team release, hosted on the VOSpace file system operated by CANFAR[4]. This release contains the same data that is delivered to the Joint ALMA Observatory (JAO) and the Canadian Astronomy Data Centre (CADC) archives. The content of the files in the team release is identical to the archive deliveries, though some metadata may be different to conform to archive guidelines.

The data delivered to JAO and released via the ALMA archive needs to comply with observatory policy. The JAO release thus includes ONLY the data from the pilot programs and the PHANGS-ALMA Large Program itself. Data released here and via CADC will contain some targets from archival programs that are imaged using the PHANGS imaging pipeline.

3. ACKNOWLEDGMENTS

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

Please first cite both the PHANGS-ALMA survey paper: A. K. Leroy, E. Schinnerer et al. ApJS, subm., and the PHANGS-ALMA data pipeline paper: A. K. Leroy et al., ApJS, accepted.

Please also add the two following acknowledgments

- This paper makes use of the following ALMA data:
  ADS/JAO.ALMA\#2012.1.00650.S,
  ADS/JAO.ALMA\#2013.1.00803.S,
  ADS/JAO.ALMA\#2013.1.01161.S,
  ADS/JAO.ALMA\#2015.1.00121.S,
  ADS/JAO.ALMA\#2015.1.00782.S,
  ADS/JAO.ALMA\#2015.1.00925.S,

[4] https://canfar.net/

ADS/JAO.ALMA\#2015.1.00956.S,
ADS/JAO.ALMA\#2016.1.00386.S,
ADS/JAO.ALMA\#2017.1.00392.S,
ADS/JAO.ALMA\#2017.1.00766.S,
ADS/JAO.ALMA\#2017.1.00886.L,
ADS/JAO.ALMA\#2018.1.00484.S,
ADS/JAO.ALMA\#2018.1.01321.S,
ADS/JAO.ALMA\#2018.1.01651.S,
ADS/JAO.ALMA\#2018.A.00062.S,
ADS/JAO.ALMA\#2019.1.01235.S,
ADS/JAO.ALMA\#2019.2.00129.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\#2013.1.00803.S : PI Espada (archival), NGC 5128/Centaurus A
ADS/JAO.ALMA\#2013.1.01161.S : PI Sakamoto (archival), NGC 1365, NGC 5236 (M83)
ADS/JAO.ALMA\#2015.1.00121.S : PI Sakamoto (archival), NGC 5236 (M83)
ADS/JAO.ALMA\#2015.1.00782.S : PI Johnson (archival), NGC 7793
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\#2016.1.00386.S : PI Sakamoto (archival), NGC 5236 (M83)
ADS/JAO.ALMA\#2017.1.00392.S : PI Blanc (pilot follow up), 9 galaxies
ADS/JAO.ALMA\#2017.1.00766.S : PI Chevance (early type extension), 7 galaxies,
ADS/JAO.ALMA\#2017.1.00886.L : PI Schinnerer (large program), 54 galaxies,
ADS/JAO.ALMA\#2018.1.00484.S : PI Chevance (early type extension), 7 galaxies,
ADS/JAO.ALMA\#2018.1.01321.S : PI Faesi (very nearby extension), 3 galaxies,
ADS/JAO.ALMA\#2018.1.01651.S : PI Leroy (pilot follow up), 9 galaxies,
ADS/JAO.ALMA\#2018.A.00062.S : PI Faesi (very nearby extension), 5 galaxies,
ADS/JAO.ALMA\#2019.1.01235.S : PI Faesi (very nearby extension), 5 galaxies,
ADS/JAO.ALMA\#2019.2.00129.S : PI Querejeta (extension), NGC 1068,

4. HELP AND FEEDBACK

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

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

5. FIRST LOOK

For a quick visual overview of the available data products for each galaxy, we refer users to the PHANGS project website[5]. It contains a visual ATLAS of the galaxies either as a google directory of images or in a PDF document.

These pages include snapshots of the available maps and a summary of the key data metrics (achieved resolution, sensitivity etc.) for each PHANGS ALMA target.

6. WHAT IS INCLUDED IN THIS RELEASE

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[6])

This distribution contains a single gzipped tar file containing all the PHANGS archive files (`phangs-alma-v4p0.tar.gz`), the PHANGS sample table (`phangs_sample_table_v1p6.fits`) and two directory trees that present different organization of the same data. The `by_galaxy` directory tree contains a single folder for each galaxy in the release. That folder holds all the PHANGS products for that galaxy. The `by_product` directory tree organizes the PHANGS data by a specific product (e.g., cube, strict_mask, mom0; see below for more details). The `by_galaxy` and `by_product` trees are redundant: the same data appears in separate places to facilitate downloading subsets of the data.

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 the two different signal identification schemes. Maps of the uncertainties associated with the two-dimensional emission maps are also distributed.

The two-dimensional maps are delivered at the native resolution for each cube and then also at a series of fixed angular resolutions: 2”, 7.5”, 11”, and 15”. The angular resolution is given by the BMAJ/BMIN keywords in the FITS header.

[5] http://phangs.org/
[6] https://docs.astropy.org/en/stable/io/fits/

7. RELEASE CONTENTS

7.1 DATA PRODUCT TYPES

Data cubes (`cubes`)
We provide the cubes at their native round-beam resolution in equatorial J2000 coordinates and the LSRK velocity reference frame. The cubes are primary beam corrected and converted to main beam brightness temperature units (Kelvin).

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., 2021, accepted). Use the strict maps if you want low noise, can live with some incompleteness, and want a very reproducible calculation.

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. Use the broad maps if you want high completeness and can live with some additional noise. 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.

7.2 FILENAME CONVENTION

The filenames of PHANGS-ALMA data products follow a common convention :
galaxy_arraycombination_linetype[_resolution][_masktype][_maptype].fits

- *galaxy* : we use NGC numbers for the names of PHANGS ALMA targets whenever possible. The remaining galaxies are identified using their IC number, or another widely-used name (e.g. Circinus).

- *arraycombination* : for this release, we deliver data products constructed from the 12m+7m or 7m interferometric data feathered with the total power data. We indicate these as “12m+7m+tp” and “7m+tp” respectively in the filenames.

- *linetype* : for this release, we deliver CO(2-1) data only, which we indicate as “co21” in the filenames.

- *resolution* : Due to uncertainties in galaxy distance, we adopt a 10% tolerance in resolution. This means, for example, that data with a resolution between 75pc and 85pc according to the PHANGS Sample Table v1.6 distances will still be labelled 80pc. For native resolution data products, no resolution is specified in the filename. The delivery includes data products at fixed angular and linear resolution. The fixed angular resolution products are indicated by adding “as” (i.e. arcseconds) to the quoted resolution in the filename, while the fixed linear resolution products are indicated with “pc” (i.e. parsecs). Regardless of the filename, which can be approximate, the beam recorded in the BMAJ and BMIN keywords is understood to be accurate and correct.

- *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’ to the maptype, e.g. ‘emom0’ indicates the error associated with the moment-0 (‘mom0’) map.

7.3 AVAILABLE DATA PRODUCTS

- Emission cubes
    - Example filename: ngc0628_12m+7m+tp_co21.fits
    - Unit: K
    - Short description: CO(2-1) spectral line cubes

- Noise cubes
    - Example filename: ngc0628_12m+7m+tp_500pc_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. Noise maps are delivered at all available resolutions.

- Signal identification masks (cubes)
    - Example filename: ngc0628_12m+7m_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.

- Integrated intensity (moment-0) and associated uncertainty
    - Example filename: ngc0628_12m+7m+tp_co21_broad_mom0.fits
    - Unit: K*km/s.
    - Filename map type: mom0
    - Short description: This is 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: ngc0628_12m+7m+tp_co21_150pc_broad_tpeak.fits
    - Example filename: ngc0628_12m+7m+tp_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 tpeak12p5kms, 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: ngc0628_12m+7m+tp_co21_strict_mom1.fits
    - Example filename: ngc0628_12m+7m_co21_15as_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: ngc0628_12m+7m+tp_co21_2as_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: ngc0628_12m+7m+tp_co21_60pc_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).

7.4 DIRECTORY ORGANIZATION

- PHANGS-ALMA
    - phangs-alma-v4p0.tar.gz
    - phangs_sample_table_v1p6.fits
    - by_galaxy
        - ic5273
            - ic5273_12m+7m+tp_co21.fits
            - ic5273_12m+7m+tp_co21_2as.fits
            - ic5273_12m+7m+tp_co21_strict_mom0.fits
            - ic5273_12m+7m+tp_co21_2as_strict_mom0.fits
            - ...
        - ic5332
        - ...
    - by_product

- cubes
  - ic5273_12m+7m+tp_co21.fits
  - ic5332_12m+7m+tp_co21.fits
  - …
- strict_mom0
  - ic5273_12m+7m+tp_co21_strict_mom0.fits
  - ic5332_12m+7m+tp_co21_strict_mom0.fits
  - …
- strict_mom1
  - ...
- ...

8. KNOWN ISSUES IN THIS RELEASE

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@phangs.groups.io .

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.

- The first and last channel of some of the ^{12}CO(2-1) 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.

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: NGC 247 (in the extended sample), NGC 1313 (also in the extended sample, but see below), IC 5332, NGC 1809, NGC 2283, NGC 3239, NGC 4496A, NGC 4571, NGC 5042, NGC 5068, and NGC 7793 (in the extended sample). 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.

- NGC 1313 in the extended sample has only one half of the galaxy currently imaged due to problems with the imaging.

- NGC 5236 (M83) has high quality maps but still shows more visible low-level artifacts in the imaging (“bowling”) compared to the other data and sometimes poorer image quality where bright emission reaches the edge of the mosaics. This galaxy is among the most likely to improve with re-imaging.