Draft version October 21, 2018
Preprint typeset using LaTeX style emulateapj v. 12/16/11

TITLE

Marat Musin [1 2], Haojing Yan [2],

Draft version October 21, 2018

ABSTRACT

Keywords: infrared: galaxies — submillimeter: galaxies — galaxies: starburst — methods: data analysis

1. INTRODUCTION

1.1. *Lilly-Madau formalism*

1.2. *SED fitting as a standard technique of mass and redshift estimation*

SED fitting is now a standard technique of deriving stellar mass and photometric redshifts for a large set of galaxies. In this method multi-band photometry for a given galaxy is fitted to a series of a templates predicted by a certain stellar population synthesis (SPS) model. The best-fit template gives the parameters of the galaxy, including its redshift and mass. Historically, SPS models were using restframe optical photometry. One caveat is the degeneracy between the dust extinction and age of the stellar population, as both make the color of galaxy red, i.e. galaxy can be red because it is intrinsically red with no young massive star and ongoing star-formation, or it can be very dusty, or it can be metal-rich and metals effectively absorb light in the bluer bands. Solution to this is to implement restframe near-IR where light suffers much less extinction (comparing to restframe UV and optical) and thus the degeneracy can be broken. We aim to build the largest sample of galaxies with optical and near-IR photometry over a large sky area. The natural choice for us then is to use optical Sloan Digital Sky Survey (SDSS) and IR all-sky data from Wide-Field Infrared Survey Explorer (WISE).

1.3. *problems associated with construction of the catalog*

Blending, poor spatial resolution in IR our method – template fitting

1.4. *Goal of this paper*

In this paper we present our technique for construction a catalog of galaxies with reliable SED data in optical and near-IR in Stripe 82 field. We discuss data selection, sources identification in different bands and problems associated with it.

2. DATA DESCRIPTION

2.1. *SDSS and Stripe 82*

The imaging component of the SDSS, which was done in five broad bands ($u'g'r'i'z'$), has covered 14,555 deg^2. In most area, the SDSS only scanned for one pass at an exposure time of 53.9 seconds per band, and thus is rather shallow (for example, the r'-band 5 σ limiting magnitude is 22.2 mag). For this reason, in most cases the SDSS can only probe the normal galaxy population up to $z \approx 0.4$. However, the Stripe 82 region, which is a long stripe along the equator that spans $20^h < RA < 4^h$ and $-1.26^o < Dec < 1.26^o$, totalling in ≈ 300 deg^2 is the exception. It was repeatedly scanned ($\sim$ 70-90 times, depending on RA) for calibration purpose during the survey (Adelman-McCarthy et al. 2007), and thus the combined scans can reach much better sensitivities.

A number of teams have created deep Stripe 82 stacks and made them available to public. The first such stacks were produced by Annis et al. (2014) based on the data obtained up to December 2005 (20-35 runs), which achieved 1-2 magnitude deeper limits than the single-pass SDSS images. Several other teams (e.g., Jiang et al. 2009; Huff et al. 2014) produced different stacks using different procedures to optimize the image qualities.

Jiang et al. (2014; hereafter J14) released a new version of stacks using only the images that were taken under the best weather conditions. These stacks are ~ 0.2 mag deeper than those produced by Annis et al. (2014), reaching 5 σ limits of 23.9, 25.1, 24.6, 24.1, 22.8 mag in $u'g'r'i'z'$, respectively, and also have better PSF characteristics. We adopt these stacks in our work.

2.2. *Structure of SDSS Stripe 82 files*

We use description from J14 to present the structure of optical data. An SDSS run (strip) consists of six parallel scanlines, identified by camera columns (Figure ??). The scanlines are 13.5 arcmin wide, with gaps of roughly the same width, so two interleaving strips make a stripe that consists of total 12 scanlines (columns).

The size of each co-added SDSS image is 2854 x 2048 pixels, or roughly 18.8' x 13.5' (RA x Dec), with a pixel size of $0.396''$ and an average full width at half maximum (FWHM) of $\sim 1.5''$ in u-band, $\sim 1.3''$ in g-band, and $\sim 1''$ in r-, i-, and z-bands. In total there are 401 SDSS images in each column and overall $12 \cdot 401 \cdot 5 = 24,060$ SDSS images in all 5 bands. Each SDSS image has a corresponding weight.fits image, that records relative weights at individual pixels.

2.3. *WISE and unWISE*

WISE (Wright et al. 2010) is a near-to-mid IR space telescope launched in 2009 and has performed an all-sky imaging survey in four bands at 3.4, 4.6, 12, and 22 μm (denoted as W1, W2, W3, and W4, respectively). During its original mission phase from 2010 January 7 to

[1] Chinese Academy of Sciences South America Center for Astronomy (CASSACA), National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
[2] Department of Physics & Astronomy, University of Missouri, Columbia, MO 65211, USA

2010 August 6 (the "4-band Cryogenic" phase), WISE surveyed the entire sky 1.2 times in all four bands simultaneously until the solid hydrogen coolant in the outer cryogen tank was depleted. It then entered the "3-band Cryogenic" phase for the next 54 days, during which time it mapped an additional 30% of the sky in W1, W2 and W3. When the coolant in the inner tank was also depleted by 2010 September, only W1 and W2 are operational. The NEOWISE project took over the mission on 2010 October 1 and brought it into the four-month "Post-Cryo" phase to survey the sky in these two bands for near-earth objects until 2011 February 1 (see Mainzer et al. 2014). The telescope was then put into hibernation for the next 35 months as the funding stopped. The extended NEOWISE project reactivated it in 2013 December to continue the two-band observations ("NEOWISE Reactivation") through today.

The WISE team made three data releases separately for the 4-band Cryogenic, the 3-band Cryogenic and the NEOWISE Post-Cryo phases in 2012 March, 2012 June and 2013 May, respectively. To take the advantage of these repeated observations, the WISE team also made the "AllWISE Data Release" in 2013 November by combining all the WISE data available till then (see Cutri et al. 2013; for details). The included image products, known as the "Atlas Images" reach the nominal 5 σ limits of 0.054, 0.071, 0.73, and 5.0 mJy in the four bands, respectively.

To optimize the detection of isolated sources, the WISE team has been using a special treatment when combining images, namely, the single-exposure images are convolved with the individual point spread function (PSF) during the stacking process. However, this operation has the drawback that it reduces the spatial resolution of the final stacks, which is not desirable in many applications. To deal with this problem, Lang (2014; hereafter L14) reprocessed all the WISE images independently without the PSF convolutions, and produced the stacks that preserve the original WISE spatial resolutions. These image products of L14, dubbed as the "unWISE" images, have the PSF full-width at half maximum (FWHM) values of $6''$ in W1, W2 and W3 and $12''$ in W4. We use these unWISE W1 and W2 images for this work.

2.4. *Structure of unWISE files - once again maybe I need to omit it in the paper*

The unWISE coadds are on the same tile centers as the WISE tiles with 18,240 images per band, 1.56 x 1.56 degrees each. The tiles are named by their RA, Dec center: tile "0591p530" is at RA = 59.1, Dec = +53.0 degrees; i.e., the first four digits of the tile name is $int(RA \cdot 10)$, then "p" for +Dec and "m" for -Dec, then three digits of $int(abs(Dec) \cdot 10)$. For each tile and band W1-w4, several images are produced, we shall list only the ones that we make use of:

– unwise-0000p000-W1-img-m.fits - "Masked" image, 2048 x 2048 pixels, TAN projected at 2.75"/pixel. Background-subtracted, in units of "Vega nanomaggies" per pixel: $mag = -2.5 \cdot (log_{10}(flux) - 9)$. This is the science image, the word "masked" means that some pixels have no unmasked pixels and no measurement at all: pixel value 0 and infinite uncertainty.

– unwise-0000p000-W1-std-m.fits - Sample standard deviation (scatter) of the individual-exposure pixels contributing to this coadd pixel.

Three unWISE images centered at the same RA cover the whole width of Stripe 82 in Dec (-1.26 δ ¡ +1.26). We shall call three such unWISE images a frame. There may be up to 72 SDSS images within one frame.

3. OVERVIEW OF METHODS FOR ANALYSIS

The most critical factor in SED fitting is consistent photometry in the involved bands, i.e., the photometry should include the same fraction of light across all bands so that the colors are defined in a consistent manner. This is challenging in our case because the spatial resolutions of WISE are at least 6× worse than that of the SDSS. For this reason, the objects detected in WISE often suffer blending. Even for relatively isolated WISE sources, the photometric apertures appropriate for the (low resolution) WISE images cannot guarantee the same fraction of light being included as what is done in the (high resolution) SDSS images. Such a systematic offset, which is different for every galaxy, severely skews the SED fitting.

To best address this problem, we opt to use the T-PHOT software, which recently emerged as a robust and flexible tool to perform "template fitting". The basic idea is to use a high-resolution image (here an image from the SDSS) as the prior to build the morphological template of the source under question, convolve this template with the PSF of the low-resolution image (here the corresponding image from the unWISE), and fit this degraded template to the low-resolution image to obtain the total flux that is within the aperture as defined by the high-resolution image. In this way, we get reliable color information (i.e., flux ratio) in the most consistent manner.

While T-PHOT is much more user-friendly as compared to its predecessors, running this software is still non-trivial. It not only requires careful tuning of parameters but also several tedious preparatory steps with both the high- and the low-resolution images. Here we detail our procedures.

3.1. *Initial preparation of unWISE and SDSS images*

T-PHOT requires that the low- and the high-resolution images have the same orientation and the same World Coordinate System (WCS) reference position, the latter of which is defined by the FITS keywords (CRVAL1, CRVAL2). It also requires that their pixel scale ratio must be an integer. To meet these prerequisites, we carried out the following procedures utilizing the SWarp software (Bertin et al. 2002), which can subsample or bin an image to any pixel scale and then re-project to an arbitrary orientation at any tangential point.

We first rescaled the unWISE images from $2.750''$/pix to $2.772''$/pix. As the scale of an SDSS image is $0.396''$/pix, this makes the ratio of their pixel scales an integer ($2.772/0.396 = 7$). We kept the same orientation, which is always North-up and East-left, and the same reference position for each unWISE image. This process was done for both the W1 and the W2 unWISE images, which are always aligned.

For a given SDSS image, we oriented it to North-up and East-left, and re-projected it at the tangential point as defined by the reference position of the unWISE images that it lies within. In other words, the FITS keywords (CRVAL1, CRVAL2) of the re-projected SDSS image is the same as those of the unWISE images. As an unWISE image covers much larger area and thus encompasses multiple SDSS footprints, the tangential projection point of a reprojected SDSS image is often outside of its coverage. In the extreme cases, it can be as far as 0.7°, outside of the image itself.

About 30% of the SDSS images lie across two adjacent unWISE fields, and therefore need to be treated separately. If one SDSS image has more than 60 arcmin^2 belonging to adjacent unWISE fields, such image is duplicated, and each copy is reprocessed with respect to the appropriate unWISE field as described above. This increases the number of SDSS images from 4,812 to 5,556 per band. Regions for which overlap between unWISE and SDSS images is less than 60 arcmin^2 are not processed. This operation excludes 11.34 deg^2 from the total area for which catalog is constructed.

We note that the above procedures were done for both the science images and the standard deviation (for the unWISE) or the weight (for the SDSS) images. After the subsampling and re-projection, the standard deviation or the weight value per pixel no longer preserves the absolute scale. In other words, the value of a given pixel on an unWISE (SDSS) reprojected standard deviation (weight) image no longer reflects the true standard deviation (weight) on that pixel. Fortunately, this does not affect the performance of T-PHOT, as it only uses these values in a relative sense (i.e., the absolute scale does not matter). However, it will affect the final errors that T-PHOT report, which we will remedy separately (see Section x.x below).

We also note that one special treatment needs to be done for the saturated pixels in the unWISE standard deviation images. They are all assigned zero standard deviation in the unWISE release, which is invalid for T-PHOT. We therefore use the IRAF/imcalc task to set such values to "9999" before reprocessing.

3.2. *Input SDSS source catalog*

J14 produced object catalogs from their stacked images using `SExtractor` (Bertin & Arnouts 1996). While in general these catalogs could be used in our project, there are several caveats. The catalogs are not cleaned out from the duplicate sources in the overlap area between two adjacent images and two adjacent scanlines; catalogs are not matched between the bands and thus have different number of sources for each band; object detection threshold was set too high (DETECT_THRESH = 2) and many faint objects were excluded; bright and saturated objects have multiple detections, none of which is centered on the source. For these reasons new catalogs will be constructed.

3.2.1. *Rationale*

Due to the great depth, roughly 24.6 AB and $\approx 1''$ PSF FWHM r-band was chosen as a reference band. Centroids, morphology, and other non-amplitude parameters of the detected sources are then fixed to the values from this reference band. Sensitivity, FWHM and local background varies in SDSS from band to band, so neither aperture, nor Kron flexible elliptical aperture (Kron 1980) can produce consistent colors in 5 optical bands, which is crucial for the subsequent SED fitting.

To remedy this problem, all individual images in $g'r'i'z'$ bands are convolved to the PSF of the corresponding image in the band with the worst PSF FWHM, namely u-band. For the price of losing some sources due to blending in the reference band, matched to the u-band, we extract more robust fluxes. Matching one images in one band to the PSF of the image in the different band requires knowledge of the kernel - a PSF matching function between two images. The PSF of SDSS images varies in RA (corresponds to image rows), due to atmospheric fluctuations and, as a result, different seeing. It also varied in DEC (corresponds to columns), due to camera optics and different airmass. Thus there is no universal PSF and individual PSF must be built for each SDSS image. Construction of the 27,780 PSFs in Stripe 82 (5,556 per band) is the most time-consuming part of the project.

3.2.2. *PSF matching*

Our strategy for the PSF construction can be described as follows. PSF is built as a combination of selected point-like sources on the image. Generally, large amount of such sources is needed to build a correct PSF that is independent of small variations in profiles of the sources *I know it should be said, but I am not sure this is the best way.* The first selection of point sources that contribute to the PSF model is performed using a comparison of the core magnitude vs. total magnitude. The ratio between these two quantities represents a concentration of the source. SExtractor MAG_APER and MAG_BEST were used to quantify core and total magnitudes respectively. Size of the aperture depends on the particular band, but in general is $\approx$ X$''$. The second selection involved magnitude cuts (saturated and faint sources were removed) and use of the stellarity index. The later was estimated using SExtractor CLASS_STAR parameter. Sources with CLASS_STAR<XX were rejected. The last selection only left the source in the final sample if it is not close to the edge of the image and does not have another nearby source within XX$''$.

Finally, each SDSS image has 20 to 160 point-like sources that contribute to the PSF model. Lower value appears in a few images when there are not enough bright point sources (mostly in the less sensitive u-band), while the cut in the maximum number of the point sources was performed in order to have stable performance of the `IRAF/psf` task. *I know it sounds ugly, but I don't want to write the IRAF sometimes crashes when you supply more than 160 stars.* Visual inspection is given to all constructed PSFs. If it shows some features, such as elongation, gradient of the background due to the nearby (within up to several arcmin) source, or faint non-detected blended source in the vicinity of the main profile, then manual selection of the point sources and reconstruction of the PSF is performed. `IRAF/seepsf` is used to take the PSF, computed by the `IRAF/psf` and build a 21x21 pixel output FITS image, consisting of the sum of the analytic function and the residual. All PSF images are subsequently normalized to the unity total counts using `wcstools/sumpix` (Mink 1998) and `IRAF/imarith` tasks.

Kernels are constructed by supplying two PSF to the `IRAF/lucy` task that uses algorithm developed by Richardson (1972) and Lucy (1974). Matching to the PSF of the u-band is performed by `IRAF/psfmatch` task that uses $g'r'i'z'$ image and a relevant kernel as an input.

3.2.3. *Catalog construction*

Catalogs are produced running `SExtractor` in the dual mode, where r-band matched to u-band ("r matched to u" for short) is used for detection and all five bands (u-band, g matched to u, r matched to u, i matched to u, z matched to u) are consequently used for photometry. Catalogs in each image are then matched by their ID.

Convolving the image with the PSF is often used for better detection of faint objects as it increases SNR of objects *I saw it in some paper, just need to find it.* As a result, even spurious detections have SNR >5 and real objects have non-realistic associated magnitude errors, which affect the choice of proper templates for SED fitting. We mitigate such problem by correcting magnitude errors for bands $g'r'i'z'$. For each image `SExtractor` was ran again in the dual mode with r matched to u as a detection and original $g'r'i'z'$ SDSS images for photometry. Two catalogs for each image (original image and matched to u-band) are stacked and the mean ratio of the original SNR to matched SNR for all sources is calculated. All magnitude errors in a matched image are then multiplied by this ratio. Besides statistical errors there are systematic error that are accounted by adding in quadrature a value of 0.04 magnitudes. Each source in each band is thus assigned with the error magnitude that was calculated using the following equation:

$$magerr\_corrected = \sqrt{0.04^2 + (\frac{1.0857}{SNR} \cdot K)^2}$$

SNR is FLUX_ISO / FLUXERR_ISO and K is correction coefficient. SNR >5 cut was applied and the catalog at this stage consists of 26,585,000 sources.

insert a figure that shows this correction

3.3. *Kernels for T-PHOT*

Template fitting requires a kernel - PSF matching function between the high resolution and the low resolution images. We used the same strategy as outlined in the section 3.2.2 and also positions of the PSF sources in r-band to construct PSF in "r matched to u" band which is used as a prior for the unWISE W1 and W2 low resolution images. Due to broader profiles after PSF matching, some PSF sources now have contamination from the neighbor sources. Such PSFs are reconstructed using IRAF/psf in the interactive mode.

Though generally space telescopes have stable PSF due to the absence of the atmosphere variations and PSF models averaged over the focal plane can be used (such approach is used in e.g. D14), we constructed individual PSF for every unWISE image in the Stripe 82 footprint. The reason behind it is that WISE Atlas Images are constructed by co-adding many single-exposure images. Because the number and relative orientation of single-exposures differ significantly between Atlas Tiles, and because the single-exposure PSFs vary with focal plane location, the PSF will be different for every Atlas Image and will vary with position on any given Atlas Image. *mostly took it from allsky supplementary materials*

There are 240 unWISE images within Stripe 82 footprint per band. 480 PSFs are constructed in a similar way to the SDSS PSFs with only one change, namely all point-like sources were selected in interactive mode using IRAF/psf. This is done to perform more robust selection of the point sources as 3 PSFs from one unWISE frame are convolved with up to 72 SDSS PSFs and thus its quality is crucial - incorrect PSF profile leads to wrong flux estimation and also creates characteristic positive and negative ring-shaped patterns in the residuals.

All unWISE PSFs are normalized to unity total counts using IRAF/imarith and sub-sampled in size by the factor of 7 using IRAF/imlintran to match the pixel scale ratio between unWISE and SDSS images. IRAF/lucy is used to convolve normalized "r matched to u" PSFs with normalized and sub-sampled unWISE PSFs to create individual kernels.

3.4. *Template fitting*

T-PHOT is a software designed to perform a precision photometry on a low-resolution images using information provided by a high-resolution images of the same field as a prior. Running T-PHOT on such a large portion of the sky with individual kernel for every pair of high- and low-resolution images is a key feature of our project.

Following recommendations from Merlin et al. (2016) two passes are performed on each pair of images. The first pass performs template fitting using provided kernel and also constructs individual kernels for each source on a given image by applying shifts in high-resolution image reference pixels along X and Y directions. These individual kernels are then used in the second T-PHOT pass. Standard pipelines are used for both passes. Single fitting is applied instead of dividing the low-resolution image on a number of individual cells. The former approach requires large amounts of computational time, but it provides the most accurate flux estimate.

One of the advantaged of T-PHOT is a large saving of computational time comparing to its forerunners, TFIT or CONVPHOT codes, but it still needs lots of CPU time. T-PHOT has to be run twice (first pass and second pass) in two unWISE bands, W1 and W2 on each of 5556 SDSS images within Stripe 82. One single pass takes $\approx$ 3 hours CPU time totaling in 66,700 CPU hours for Stripe 82. Computation is performed on the University of Missouri High-Performance Computing (HPC) cluster Lewis to process all images. *Do I need to insert a figure that shows residual here?*

Once all unWISE images within Stripe 82 footprint in W1 and W2 bands are processed by T-PHOT, output catalogs need to be matched with the optical ones. Matching is performed with STILTS code separately for every image using unique source ID as the matching parameter. Due to poorer sensitivity of the WISE as compared to SDSS, XX% and YY% of sources in bands W1 and W2 respectively were assigned with zero or negative flux. We keep such sources in the final catalog and for the SED fitting treat them as the upper limit of XX and YY AB magnitudes respectively.

check how EAZY treats non-detection

I rewrote till this phrase

4. REDSHIFTS AND MASSES

The availability of near-IR filters helps improving the z phot accuracy beyond z $\approx$ 1.3, where the 4000A break goes out of the z 0 filter and the Lyman break is not yet detectable in the u - band. (from 1809.03373.pdf)

4.1. *EAZY*

redshift determination, matching to the spectroscopic redshifts comparison of the redshift with and without unWISE data (similar results)

4.1.1. *Star-Galaxy separation*

4.1.2. *Removing duplicate sources*

Adjacent SDSS images have 25" and 28" wide overlapping regions in RA and Dec respectively. Sources that appear in such regions have to be removed from the final catalog. Internal match on the catalog is performed and duplicate sources within 1.2 asec matching radius are removed.

4.1.3. *Masking bright regions*

Objects that are extremely bright in near-IR present a number of problems to the photometry measurement because it generates a host of artifacts. These artifacts include halos (low surface brightness emission extending well beyond the PSF), diffraction spikes, horizontal stripes and residual ghosts. If not corrected, these artifacts will compromise the reliability of W1 and W2 photometry performed by TPHOT. To account for such artifacts the natural solution is to mask certain regions based on available catalogs. We used SAO star catalog [Staff, 1966] and Bright IR Stars Compilation (BIRSC) [R. Tam and C. Xu - IPAC] as a base for the masking catalog. There are around 400 stars in both catalogs that fall within Stripe 82 footprint, but our visual inspection showed that some very bright objects (stars and galaxies) that severely alters results of TPHOT are not in this list. We inspected all residual FITS images and added 300 more objects to the list that now consists of 706 objects. For stars from the catalog the masking area depends on the provided V-band magnitude and it was assigned manually based on the estimation of the halo for the added objects. Radii for masking are ranging from 50 to 500 arcsec. An example of such source with masking radius=100 asec is presented on Figure 4.8. The overall masked area is 3.871 deg2, which reduces the total sky area of the survey to 288.212 deg2. After rejection of such objects our catalog contains 9,061,068 objects and this is the final sample that we use as input to SED fitting codes that return photo z and stellar mass that we use to build GSMD.

4.2. *FAST*

5. RESULTS

6. DISCUSSION

7. SUMMARY

REFERENCES

Adelman-McCarthy, J., M.A., A., S.S., A., et al. 2007, VizieR Online Data Catalog, 2276

Annis, J., Soares-Santos, M., Strauss, M. A., et al. 2014, ApJ, 794, 120

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228

Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Tech. rep.

Huff, E. M., Hirata, C. M., Mandelbaum, R., et al. 2014, MNRAS, 440, 1296

Jiang, L., Fan, X., Bian, F., et al. 2009, AJ, 138, 305

—. 2014, ApJS, 213, 12

Kron, R. G. 1980, ApJS, 43, 305

Lang, D. 2014, AJ, 147, 108

Lucy, L. B. 1974, AJ, 79, 745

Mainzer, A., Bauer, J., Cutri, R., et al. 2014, in Lunar and Planetary Inst. Technical Report, Vol. 45, Lunar and Planetary Science Conference, 2724

Merlin, E., Bourne, N., Castellano, M., et al. 2016, Astronomy and Astrophysics, 595, A97

Mink, D. J. 1998, in Bulletin of the American Astronomical Society, Vol. 30, AAS/Division of Dynamical Astronomy Meeting, 1144

Richardson, W. H. 1972, Journal of the Optical Society of America (1917-1983), 62, 55

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

APPENDIX

APPENDIX

ACKNOWLEDGMENTS

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. Part of our data processing and analysis were done using the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC).