NASA's Giovanni and AppEEARS: Tools to Facilitate Linking People to Pixels in the Social and Health Sciences

Panel contribution to the Population-Environment Research Network Cyberseminar,
"People and Pixels Revisited"
(20-27 February 2018)
https://populationenvironmentresearch.org/cyberseminars/10516

By Alex de Sherbinin and Susana Adamo
NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN),
The Earth Institute at Columbia University, USA

The *People & Pixels* volume demonstrated the power of linking remote sensing to socioeconomic data to advance important research avenues in the social and health sciences. What may not have been so apparent to readers, but was abundantly evident to the volume's contributors, was the pain and suffering endured by these early pioneers to wrangle the data from two completely different disciplinary domains into formats that would allow for integrated analysis. On the remote sensing side there were atmospheric corrections and tiling, among other laborious tasks, and on the social science side there were field data collection and compilation of census or survey data in spatial formats for use in the common platform of the day, ARCInfo. Entire semesters of graduate student's time were taken up with the painful work of selecting (cloud free only!), downloading and processing the remote sensing imagery. The NASA-funded Socioeconomic Data and Applications Center (where we both work) sought to ease the burden of data integration by transforming census data from administrative units to grid cells, so that it could be more easily analyzed in conjunction with remote sensing and other raster data sets. But for local area studies, researchers often still needed to compile "fit for purpose" data that required a lot of manual processing to get their data to "play nice".

This paper introduces two NASA tools, Giovanni[1] and AppEEARS[2], that take much of the pain and suffering out of remote sensing data discovery, extraction, processing and visualization. Our goal is to briefly introduce these tools to researchers in the social and health sciences who would like a quick way to extract relevant remote sensing parameters for analysis in a spatial or tabular framework, or to do exploratory data analysis without having to expend a lot of time and effort. The tools are also great for data visualization, classroom teaching, and lab exercises!

Giovanni's focus is largely on remote sensing-derived atmospheric data of relatively low spatial resolution and very high temporal resolution (e.g. even hourly data). The tool uses server-side processing to process requests "on the fly", and produces very nice map and graph visualizations. AppEEAR's focus is largely on remote sensing-derived terrestrial data of moderate spatial resolution (30m to 1km) but lower temporal resolution (e.g., daily time steps or monthly composites). It also uses server-side processing, and delivers your results within minutes or hours

[1] Giovanni = GES-DISC (Goddard Earth Sciences Data and Information Services Center) Interactive Online Visualization ANd aNalysis Infrastructure
[2] AppEEARS = Application for Extracting and Exploring Analysis Ready Samples

(depending on the degree of processing) by email notification. Both tools enable analysis of time trends – Giovanni for one pixel or averaged over a larger area comprising many pixels, and AppEEARS for multiple points (geographic coordinates) or polygons (area of pixels). Table 1 provides a summary of the capabilities available in each tool.

Table 1. Comparison of Giovanni and AppEEARS

Tool	Spatial Res.	Temporal Res	Time Series Analysis	Imagery Subsetting	Area Selection	Data Viz	Download
Giovanni	Low	Hourly to daily	One pixel or area averages	Yes	Bounding box, pre-defined geographies**	Yes	Imagery*, Tabular, Images±, Charts
AppEEARS	Moderate	Daily and monthly (and soon hourly)	Multiple points or polygons	Yes	Bounding box, upload SHP or GEO-JSON	Yes	Imagery***, Tabular, Images, Charts

* GeoTIFF; ** countries, U.S. states, and major watersheds; ± PNG, KML; *** GeoTIFF, NetCDF4;

Here we provide brief introductions to the tools, beginning with Giovanni because it has a longer history (first release was in 2002), followed by AppEEARS in 2016.

Giovanni[3]

Giovanni's remote sensing assets are concentrated primarily in the areas of atmospheric composition, atmospheric dynamics, global precipitation, hydrology and solar irradiance. More than 1,850 variables are available, searchable via a key word and faceted search. Example variables of potential interest to social and health scientists include:

- Precipitation rates from the Tropical Rainfall Measurement Mission (TRMM)
- Flooding from North American Land Data Assimilation System (NLDAS) (North America only)
- Air temperature at 2m and numerous other climatic variables from NLDAS
- Aerosol Optical Depth converted to PM2.5 from the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) atmospheric reanalysis for the satellite era
- UV exposure from the OMI/Aura Surface UVB Irradiance and Erythemal Dose Daily

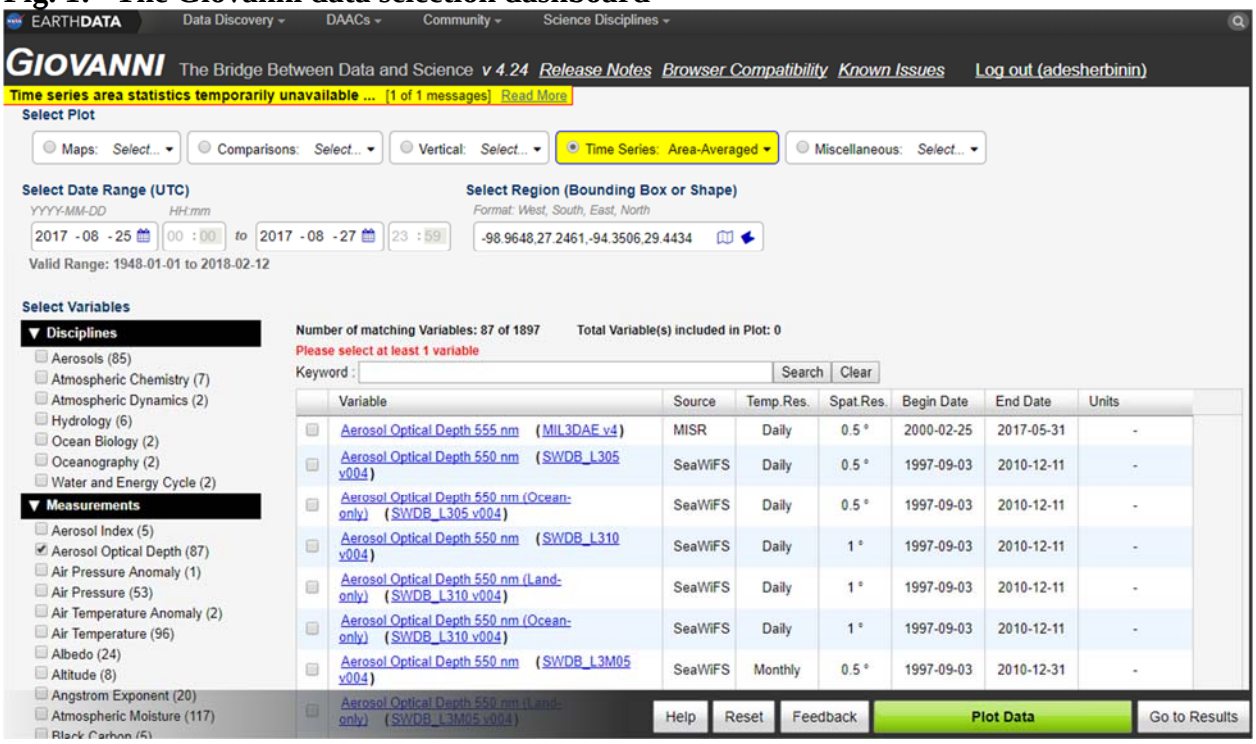
The variables vary in their spatial and temporal resolutions, as well as the time series available. For each variable the definition, source, temporal and spatial resolution, begin and end dates, and units (including choice of different units) are available in tabular format (see Fig. 1).

Mapping options include time averaging, animation, precipitation accumulation, time-averaged overlay of two data sets, and user-defined climatology. Maps can be downloaded in PNG, or in GeoTIFF or NetCDF for further analysis in a GIS environment. For time series plots and CSV downloads, options include area averaged, differences, and seasonal data (e.g. July-August for daily max temepratures over mutiple years). Visualization features include interactive map area adjustment (zoom to an area of interest in resulting maps), animation, interactive scatterplots, date

[3] Portions of this write up are based on Liu & Acker 2017

range adjustment, choice of color palette, contouring, and scaling (linear or log). The on-the-fly area adjustment feature allows a user to examine a result map interactively and in detail without replotting data.

Fig. 1: The Giovanni data selection dashboard



Example outputs can be found in Fig. 2 and 3. Figure 2a represents average hourly rainfall from 28-29 August 2017 (during Hurricane Harvey) as measured by TRMM over the Texas Coast, and Figure 2b represents the average 3-hourly rainfall rates over the same area from 25-28 August. \ Figure 3 represents the temperature at 2m above ground level in the New York metropolitan area from 1-31 July 2010. Both plots Fig. 2b and 3 can be downloaded as a CSV time series.

Fig. 2a: Rainfall over the Texas coast (mm per hour) from 25-29 August 2017

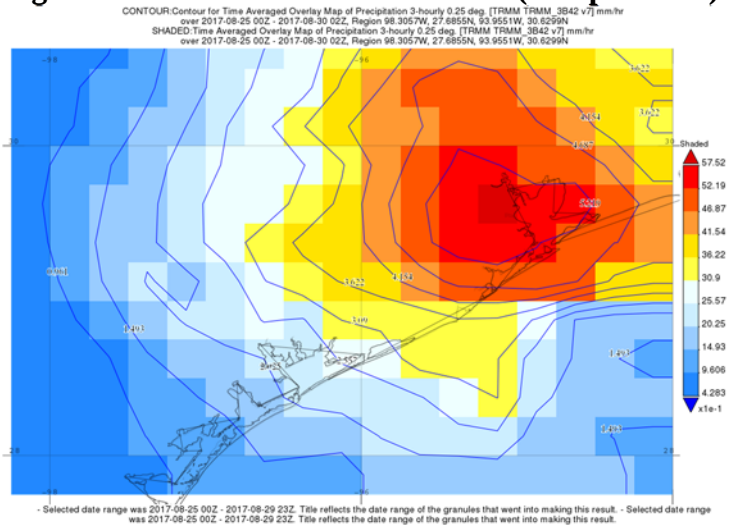


Fig. 2b: Rainfall over the Texas coast (mm per hour) from 25-29 August 2017

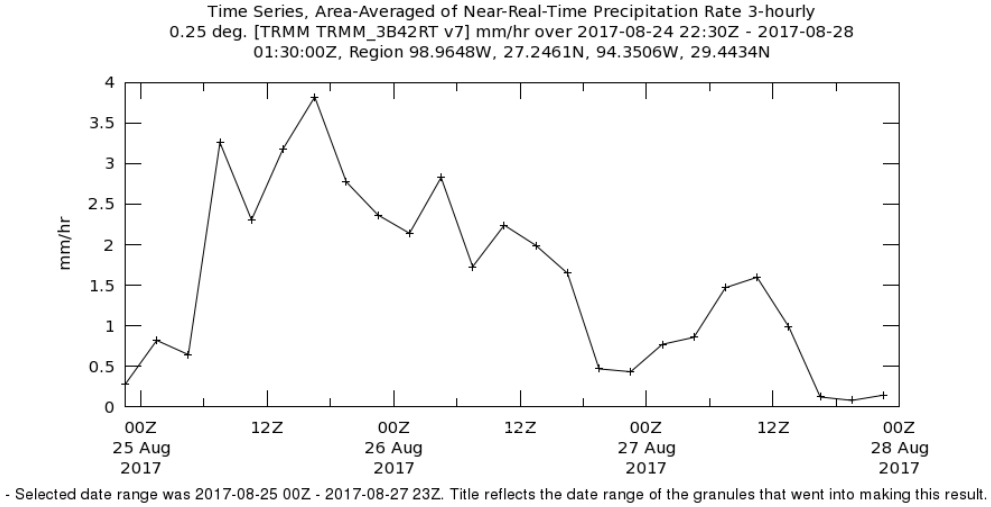
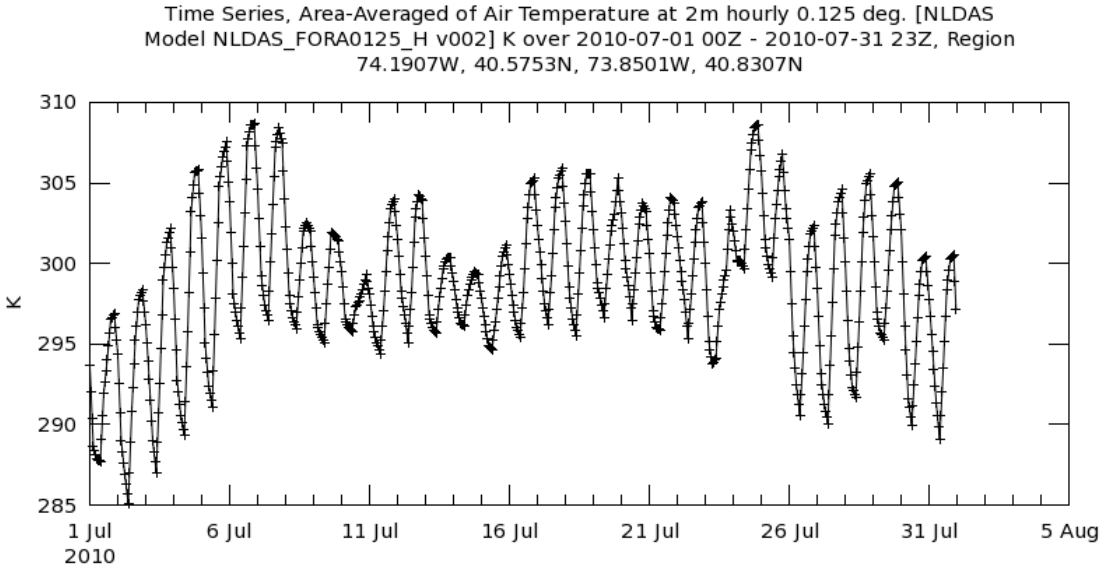


Fig. 3: Air temperature New York metro area 1-31 July 2010



While these examples represent relatively short time series, it is possible to generate far longer time series data over an area for use in conjunction with socioeconomic or health data. An example might be emergency room visits in relation to temperature (extreme heat) or PM2.5 concentrations, or exposure of a particular demographic group to repeated extremes.

One additional feature of Giovanni is persistent URLs, such that the exact query used to generate a result can be communicated via a long URL that stores all the selected parameters.

AppEEARS[4]

AppEEARS provides access to more than 100 datasets from Terra & Aqua MODIS, NASA MEaSUREs Shuttle Radar Topography Mission (SRTM v3), NASA MEaSUREs Web Enabled Landsat Data (WELD), Gridded population of the World (GPW), and NASA data products derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) Instrument.

Selected datasets of interest to social scientists include:

- Land surface temperature from Terra or Aqua MODIS
- Land cover type from MODIS (5 classifications available)
- Population counts and densities from SEDAC's Gridded Population of the World, v4
- Elevation from the Shuttle Radar Topography Mission (SRTM)
- Vegetation indices from Terra or Aqua MODIS

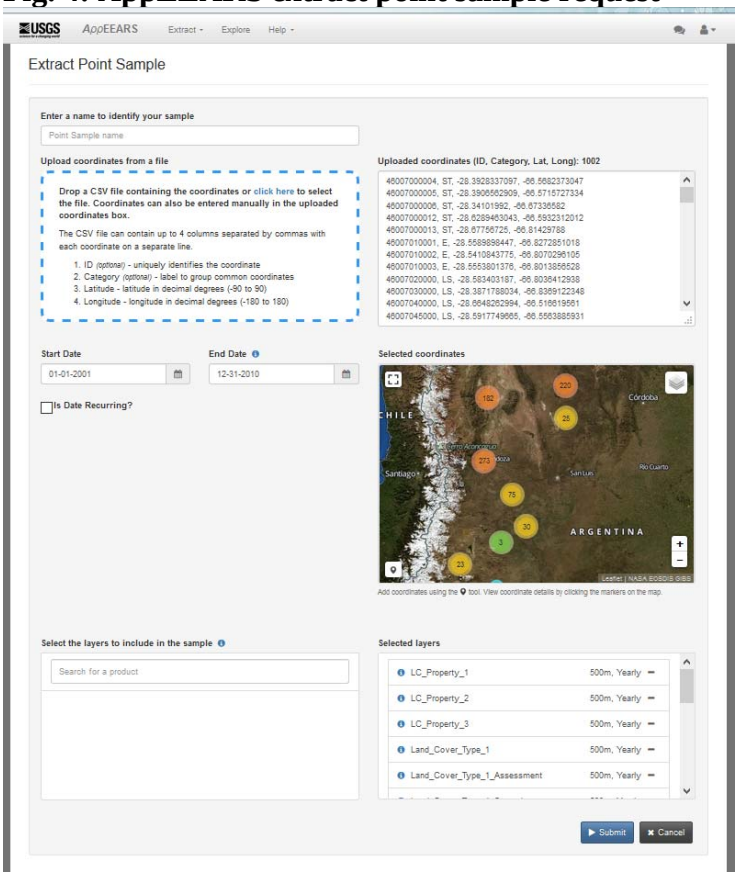
A key feature of AppEEARS is the ability to subset large geospatial datasets and to extract data for specific locations or regions around the world. Data can be extracted using an area sample via vector polygons (bounding box, Shapefile, or GeoJSON) or a point sample with geographic coordinates (latitude and longitude). In addition to data values, the user also receives quality values. The results preview includes interactive visualizations, decoded quality information, and summary statistics. The Help tab on the AppEEARS website provides step-by-step instructions for how to use the application as well as links to documentation for each geospatial dataset available.

The example below is from a point sample of settlements located in Argentina's west central drylands (Fig. 4). Each point is associated with different categories of settlements according to an Argentinian classification system (locality subdivision, agglomeration, single locality, and built structures). The objective is to add land cover class correlates (including changes over time) to population and other demographic information already in the settlements database. A research question might be, for example, when are population changes (e.g., increase or decline in the number of people and households) important enough to trigger changes in the land cover classification (e.g., from quasi-natural to urban)? The selected dataset (MCD12Q1, Version 5) is part of the Combined Terra and Aqua MODIS data collection, and provides information characterizing five different global land cover classification systems.[5] In addition, the data product provides a land-cover type assessment, and quality-control information. The results of the point sample can be downloaded as a zip file containing a CSV file with land cover types matched to the geographic points in the original sample.

[4] https://lpdaacsvc.cr.usgs.gov/appeears/

[5] https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1

Fig. 4: AppEEARS extract point sample request



After the request has been processed, the user has the option of interacting with the results online and then downloading the results in tabular format. The example below uses the layer comparison tool in the interactive window.

Fig. 5: Comparison of two classification schemes for a single settlement point, 2001-2010

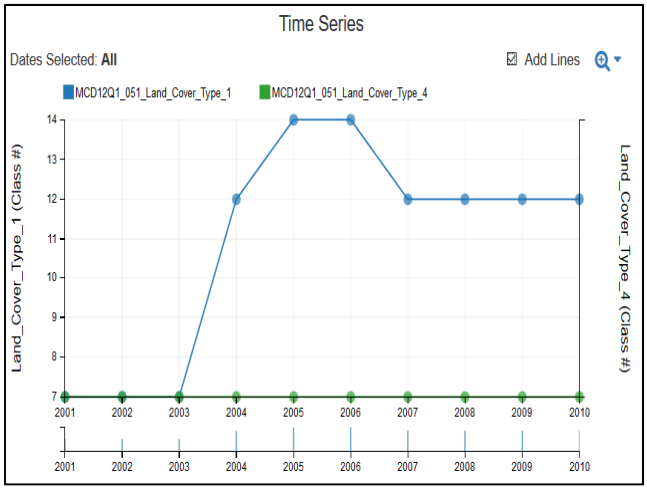
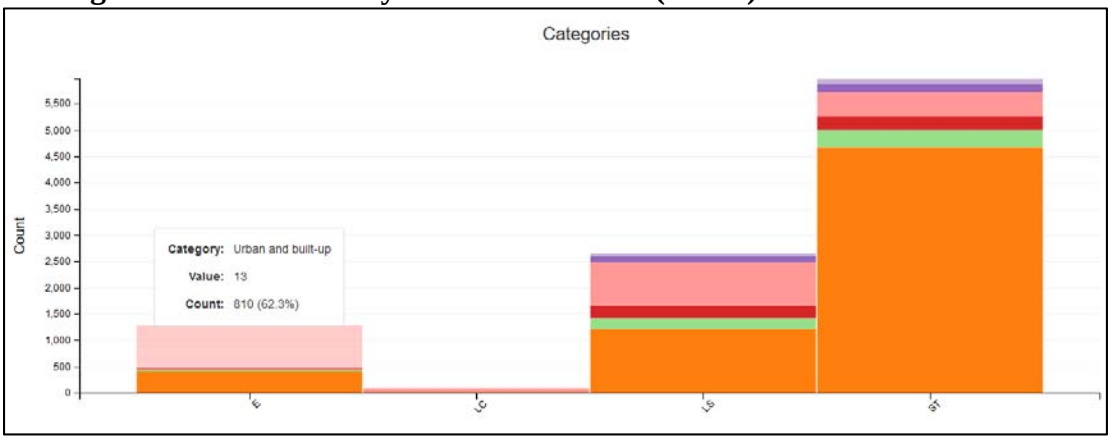


Fig. 5 displays a comparison of two different land cover classifications (Land Cover Type 1 and 4) for the settlement point identified in the map (classified as a built up place, or place where

structures have been identified), which is located toward that the northern end of the Tulum Oasis, an area of expansion of the San Juan metropolitan area into irrigated cropland. Land Cover Type 1 (LCT1) is the IGBP (International Global Biosphere Programme) global vegetation classification scheme, and Land Cover Type 4 (LCT4) is the MODIS-derived Net Primary Productivity (NPP) scheme. The two classifications differ in their assessment of the settlement point: for LCT4, it is "non-vegetated land" over the whole time-period, while for LCT1, it alternates between open shrublands, cropland and cropland/natural vegetation. A possible conclusion for this specific location is that LCT1 is more sensitive to seasonal or annual changes in vegetation than LCT4, and this information could inform data selection for a follow-on analysis.

A follow up question would be: what proportion of the settlement points are located in land cover areas identified as urban? Figure 6 displays the distribution of the point sample by categories of settlements and land cover class in LCT1. The numbers appear when mousing over the graph in the categorical overview tool in the interactive window. The graph indicates that for locality subdivisions (type E), 62.3% are classified as 'urban and built up' land cover, whereas for single localities (LS), only 30.9% are similarly classified (pink shaded area).

Fig. 6: Categories of settlement by land cover classes (LCT1)



Key: E: locality subdivision; LC: agglomeration; LS: single locality; ST: built structures

This example can easily be scaled up, for example by using the Global Rural-Urban Mapping Project (GRUMP) settlement points data set (CIESIN 2017), which includes 70,000+ settlements and associated population sizes globally, to create the point sample.

Several references in the reference section provide details about AppEEARS features and capabilities as well as use case examples. In addition, two learning/training resources are also available:

- Choosing a Data Access Tool: AppEEARS Area Sampler
  https://www.youtube.com/watch?v=6aqGfR-9ef8
- Using NASA's AppEEARS to Slice and Dice Big Earth Data
  https://lpdaac.usgs.gov/sites/default/files/public/AppEEARS_Webinar_presentation_v03_Oct2017.pdf

Acknowledgments

We thank David Meyer and Mahabaleshwara Hegde of Goddard Earth Sciences Data and Information Services Center (GESDISC) and Lindsey Harriman and Cole Krehbiel of the Land Processes DAAC (LP DAAC) for comments on earlier drafts of this paper.

References

Center for International Earth Science Information Network (CIESIN), Columbia University, CUNY Institute for Demographic Research (CIDR), International Food Policy Research Institute (IFPRI), The World Bank, and Centro Internacional de Agricultura Tropical (CIAT). 2017. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Settlement Points, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4BC3WG1.

Harriman, L. and A. Friesz. 2018. AppEEARS 2.0: New Features and Data to Further Improve User Workflow Efficiency. Available online at: https://earthdata.nasa.gov/appeears-2-0

Harriman, L. and A. de Sherbinin. 2017. From People to Pixels: Integrating Data Across the NASA DAACs. ICSU-WDS Blog. Available online at: https://www.icsu-wds.org/news/blog/from-people-to-pixels-integrating-data-across-the-nasa-daacs

Krehbiel, C.; Friesz, A.; Harriman, L.; Quenzer, R.; Impecoven, K.; Maiersperger, T.. 2016. Tools and Services for Working with Multiple Land Remote Sensing Data Products. American Geophysical Union, Fall General Assembly 2016, abstract id. IN53C-1914 http://adsabs.harvard.edu/abs/2016AGUFMIN53C1914K

Liu, Z., and J. Acker. 2017. Giovanni: The Bridge between Data and Science. *EOS*, 98, https://doi.org/10.1029/2017EO079299. Published on 24 August 2017.

NASA Land Process (LP) DAAC. 2017. Exploring GPW Population and MODIS Temperature data in AppEEARS. Available online at: https://lpdaac.usgs.gov/user_resources/data_in_action/exploring_gpw_population_and_modis_temperature_data_in_appeears

Quenzer, R. and A. Friesz. 2015. AppEEARS: Simple and Intuitive Access to Analysis Ready Data. American Geophysical Union, Fall Meeting 2015, abstract id. IN51B-1801. http://adsabs.harvard.edu/abs/2015AGUFMIN51B1801Q