Remote Sensing for Social Scientists: Tools to Facilitate the Use of Remote Sensing Data in the Social and Health Sciences Research

By Susana Adamo and Alex de Sherbinin
NASA Socioeconomic Data and Applications Center (SEDAC)
CIESIN – Columbia University

One of the roles of the NASA Socioeconomic Data and Applications Center (SEDAC) is to serve as a "bridge" between the Earth and social sciences. This guide was developed with the goal of expanding the use of NASA and other remote sensing data in the social and health sciences. Many social and health scientists are interested in including environmental variables derived from satellite remote sensing in their research, but the lack of technical expertise and the time required to process remote sensing images represent significant barriers to entry. Fortunately, there are new tools that greatly facilitate the incorporation of spatial environmental data and time series location-based estimates of environmental parameters into research projects.

This guide provides a list of web-based tools that facilitate the integration of remote sensing and socioeconomic data, and an introduction to two tools, Giovanni and AppEEARS, that have been developed by SEDAC's sister data centers (GES-DISC and LP-DAAC, respectively). Giovanni and AppEEARS greatly simplify remote sensing data discovery, extraction, and processing for 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
- Do exploratory data analysis without having to spend a lot of time and effort in data processing
- Integrate remote sensing data in visualizations, classroom teaching, and lab exercises
- Prepare visualizations of remote sensing and socioeconomic data.

The Tools

Instructions and illustrative use cases on the following two tools are provided in the pages that follow:

AppEEARS (**A**pplication for **E**xtracting and **E**xploring **A**nalysis **R**eady **S**amples) aims to significantly reduce the amount of data users needed to download and process. It enables similar datasets to be discovered more easily and provides a way to more efficiently explore results. It enables users to subset geospatial datasets using spatial, temporal, and band/layer parameters. Three layers from SEDAC's Gridded Population of the World version 4 (data quality indicators, population counts and population densities) are included among the available datasets.

Giovanni (**G**ES-DISC (Goddard Earth Sciences Data and Information Services Center) **I**nteractive **O**nline **V**isualization **AN**d a**N**alysis **I**nfrastructure) is a web application that provides a simple, intuitive way to visualize, analyze, and access Earth science remote sensing data, particularly from satellites, without having to download the data. SEDAC socioeconomic and population datasets can be easily integrated with Giovanni layers, as shown in this use case.

These additional tools also offer significant power for social and health science users interested in integrating remote sensing with socioeconomic and health data:

The **ArcGIS Living Atlas of the World** is the foremost collection of geographic information from around the globe. It includes maps, apps, and data layers to support your work. It allows to explore maps, apps, and data layers from Esri and thousands of other organizations; to combine content with the user's own data to create new maps and applications; and to share how other users are visualizing data and solving problems. A large number of SEDAC data layers – from population, to poverty, to roads – are included in the Living Atlas. Just search on the term "SEDAC" from the large search box on the home page.

Geoquery is a tool developed by AidData at William and Mary College to enables users to integrate satellite, conflict, development aid, economic, health and more subnational data, from anywhere in the world, into a single simple-to-use file compatible with Excel, STATA, or your own favorite program. Users are able to extract these data for subnational administrative units for any country. Remote sensing derived data include VIIRS and DMSP-OLS nighttime lights, MODIS land surface temperature, MODIS land cover, and two air quality metrics. A handy user user guide provides step-by-step instructions.

Google Earth Engine is a platform for petabyte-scale scientific analysis and visualization of geospatial datasets, both for public benefit and for business and government users. Earth Engine stores satellite imagery, organizes it, and makes it available for the first time for global-scale data mining. The public data archive includes historical earth imagery going back more than forty years, and new imagery is collected every day. Earth Engine also provides APIs in JavaScript and Python, as well as other tools, to enable the analysis of large datasets. SEDAC's Gridded Population of the World v4 is included as a data layer that can be used in conjunction with satellite remote sensing imagery.

IPUMS Terra provides globe-spanning data on human population characteristics, land use, land cover, and climate and makes these data interoperable by performing transformations across population microdata (individual- and household-level records), area-level data, and raster data, enabling researchers can get the variables they need in the data structure best suited to their analysis.

Worldview allows users to interactively browse global satellite imagery within hours of it being acquired. This app from NASA's EOSDIS provides the capability to interactively browse over 600 global, full-resolution satellite imagery layers and then download the underlying data. Many of the available imagery layers are updated in near real time (NRT, or within three hours of observation), essentially showing the entire Earth as it looks "right now". This supports time-critical application areas such as wildfire management, air quality measurements, and flood monitoring. Arctic and Antarctic views of several products are also available for a "full globe" perspective. SEDAC population, settlements, hazard and socioeconomic data may be added to the map viewer by clicking on the orange "Add Layers" button, then clicking on the "Science Disciplines" tab, and then selecting the "Human Dimensions" category, or by using the search box. Once a layer is selected, overlays of NRT satellite data are possible by clicking on the "layer options" from the legend, and setting opacity to 50% or lower.

A Brief Introduction to Giovanni

https://giovanni.gsfc.nasa.gov/giovanni/

Note: Registration is required for extracting, downloading and visualizing data through Giovanni. Also, remember that it is important to read the documentation, guides and metadata, and to understand any data limitations. Selected references appear at the bottom of this web page.

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.

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).

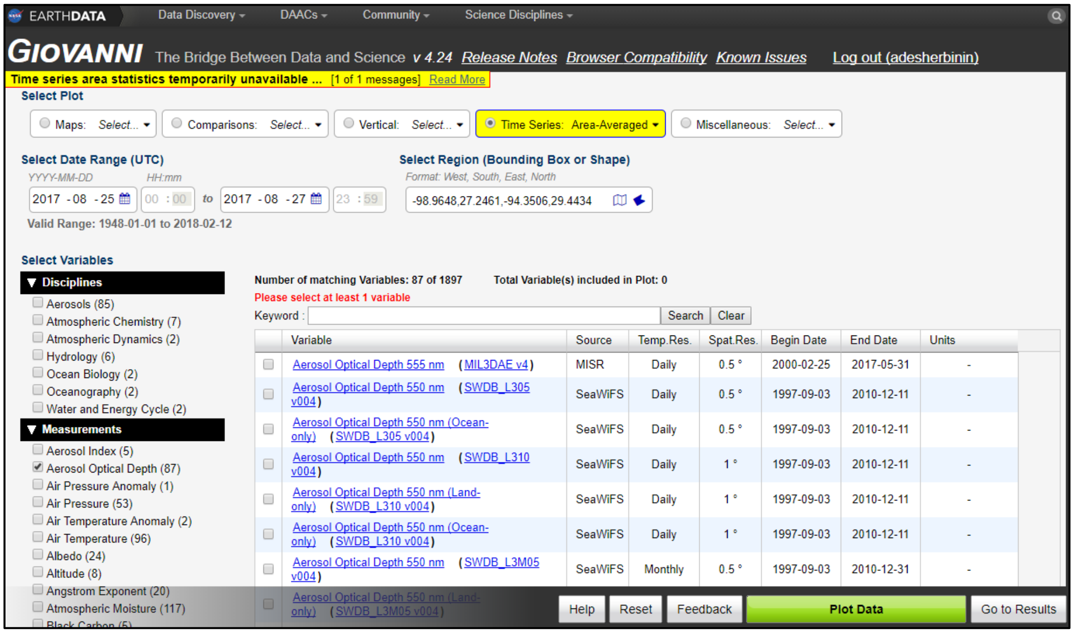


Figure 1: The Giovanni data selection dashboard

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 temperatures over multiple years).

Visualization features include interactive map area adjustment (zoom to an area of interest in resulting maps), animation, interactive scatterplots, date 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. 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.

The example outputs below represent relatively short time series, but 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.

Giovanni - Example outputs

1. Heat waves in New York

Figure 2 represents the temperature at 2m above ground level in the New York metropolitan area from 1-31 July 2010. Plots can be downloaded as a CSV time series.

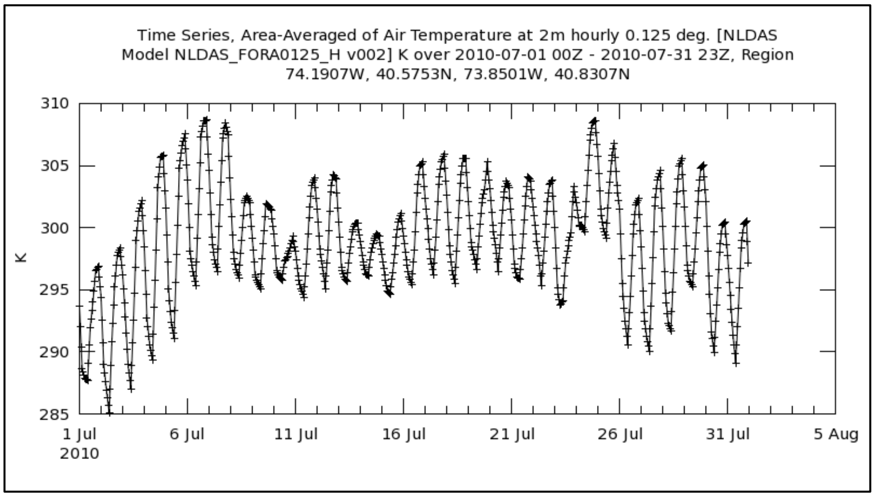


Figure 2: Air temperature New York metro area 1-31 July 2010

2. Spatial and temporal patterns of rainfall during Hurricane Harvey

Figure 3a maps average hourly rainfall from 28-29 August 2017 (during Hurricane Harvey) as measured by TRMM over the Texas Coast, showing the concentration of heavy rainfall over Houston. Figure 3b represents the average 3-hourly rainfall rates over the same area from 25-28 August, indicating that the peak rainfall occurred during 25 August.

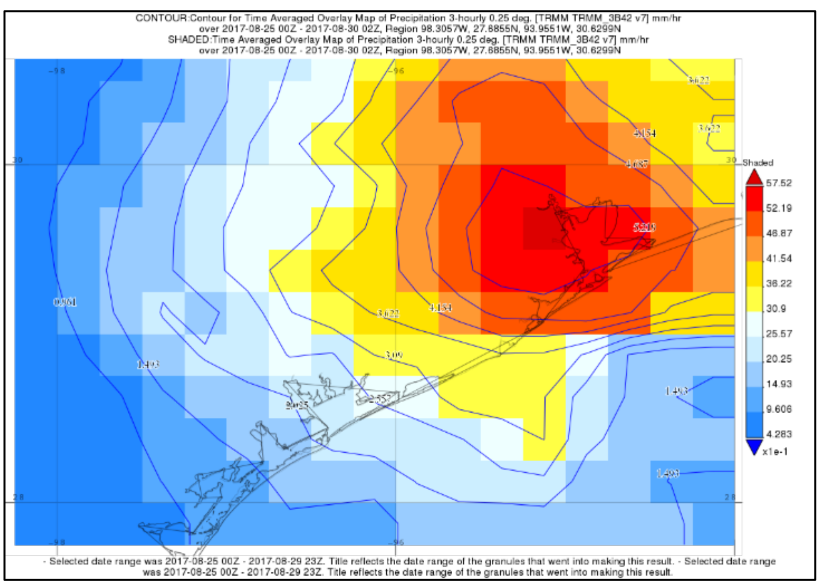


Figure 3a: Rainfall over the Texas coast (mm per hour) from 25-29 August 2017

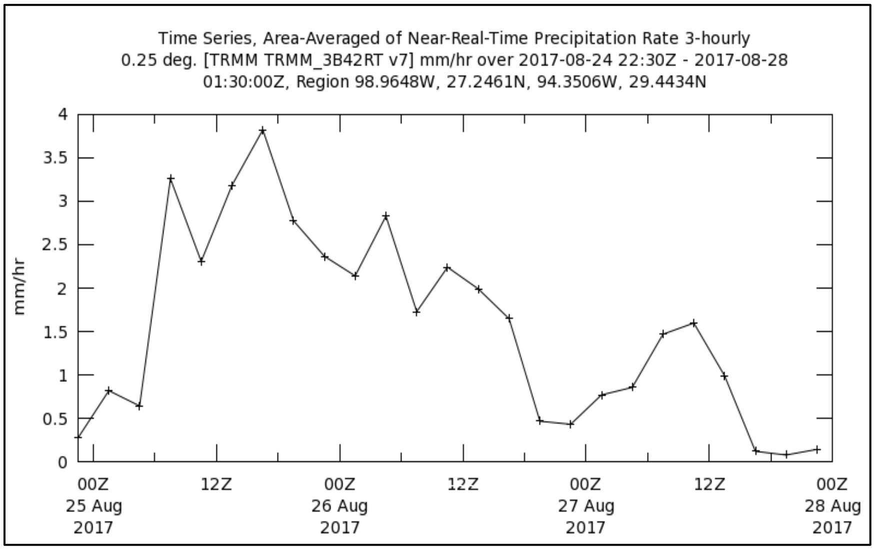


Figure 3b: Rainfall over the Texas coast (mm per hour) from 25-29 August 2017

3. Comparing the impacts of Hurricanes Irene and Sandy in the Tristate Area

Figure 4 displays the average wind speed (in meters per second, upper panels) and average rainfall (in mm per hour, lower panels) during Hurricanes Irene (left panels) and Sandy (right panels) for counties in the Tristate area and neighborhood states. High winds affected a larger area during Hurricane Sandy, especially in the coastal areas, while high rain intensities were widespread during Hurricane Irene, particularly inland.

The maps integrate outcomes from Giovanni (in tiff format) overlaid on county boundaries from SEDAC's US Census Grids (http://sedac.ciesin.columbia.edu/data/collection/usgrid).

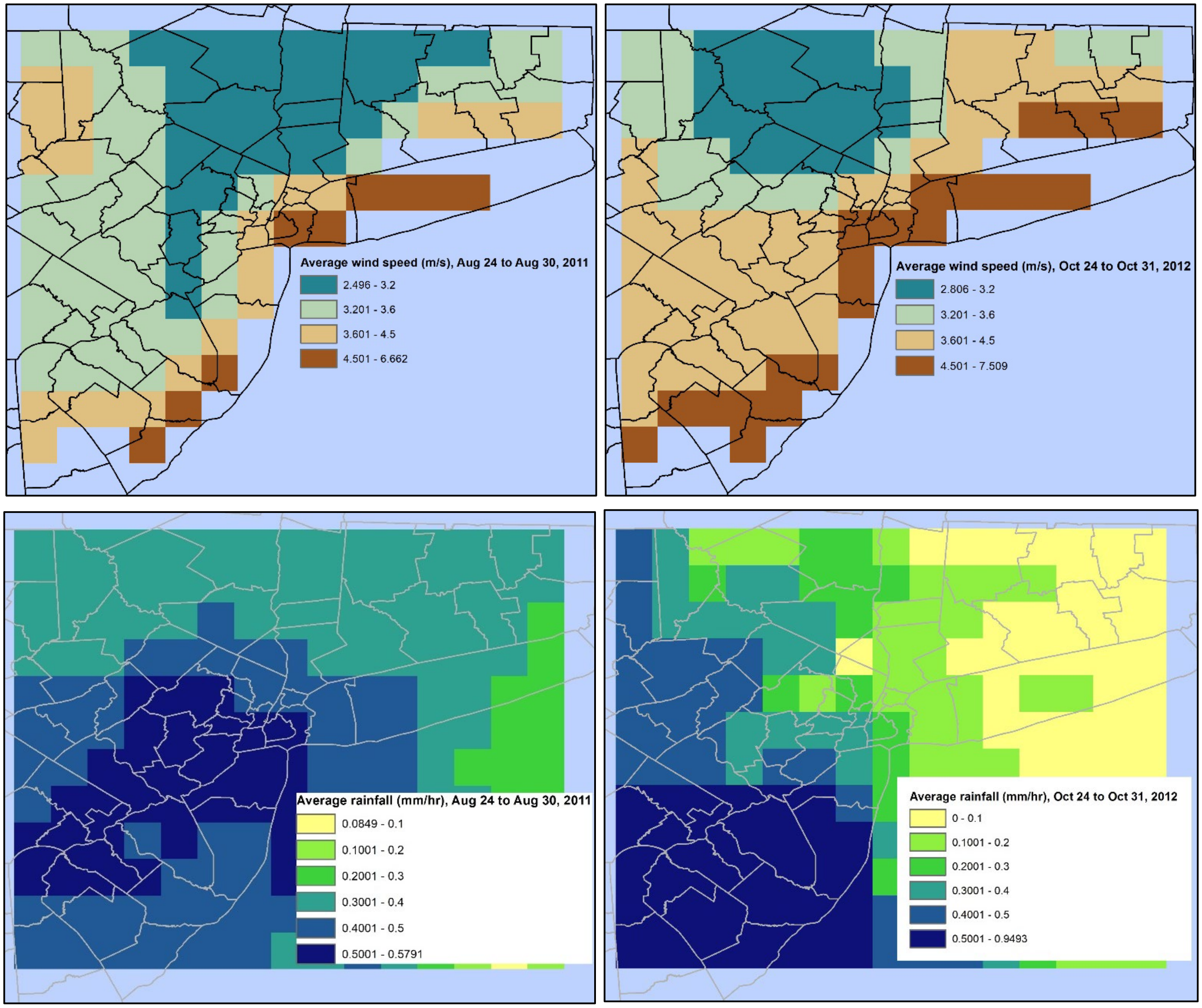


Figure 4: Wind (upper) and rain (lower) fields, Hurricanes Irene (left) and Sandy (right)

A Brief Introduction to AppEEARS

https://lpdaacsvc.cr.usgs.gov/appeears/

Note: Registration is required for extracting, downloading and visualizing data through AppEEARS. Also, remember that it is important to read the documentation, guides and metadata, and to understand any data limitations. Selected references appear at the bottom of this web page.

AppEEARSs focuses on remote sensing-derived terrestrial data of moderate spatial resolution (30m to 1km) and varied temporal resolution (e.g., daily time steps, monthly composites, or year averages). It also uses server-side processing, and delivers your results within minutes or hours (depending on the degree of processing) by email notification.

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), SEDAC's 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 (DATES)
- 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. From the Extract tab, 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). Importantly, in addition to data values, the user also receives quality values (e.g., degree of cloud cover). The results preview includes interactive visualizations, decoded quality information, and summary statistics.

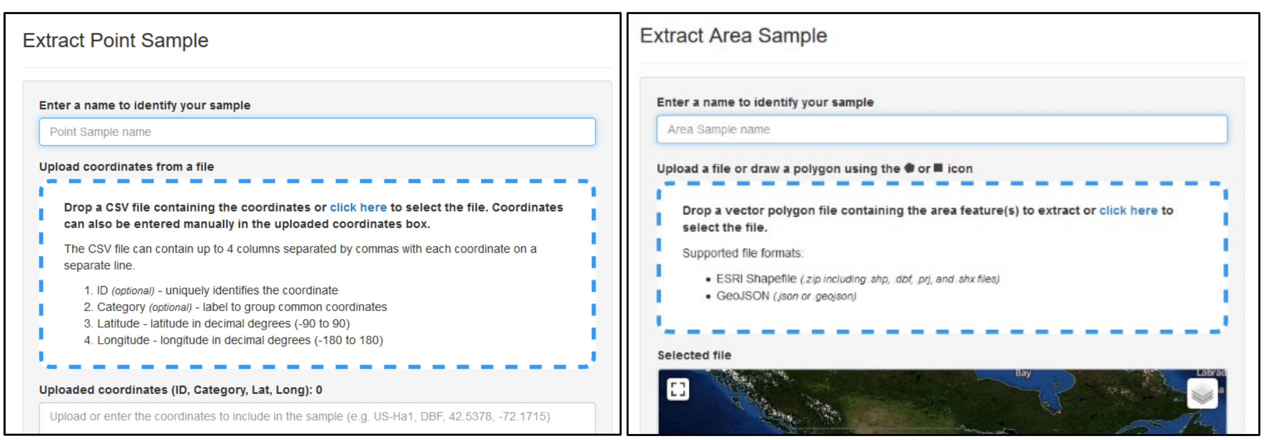


Figure 5: Data extraction interfaces for AppEEARS

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. 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

AppEEARS - Example Outputs

1. Adding population and vegetation cover information to settlement points

Cambodia saw a quick repopulation of cities and urban growth in the last decades (Figure 6). It is likely this quick urbanization process influenced vegetation cover in urban areas.

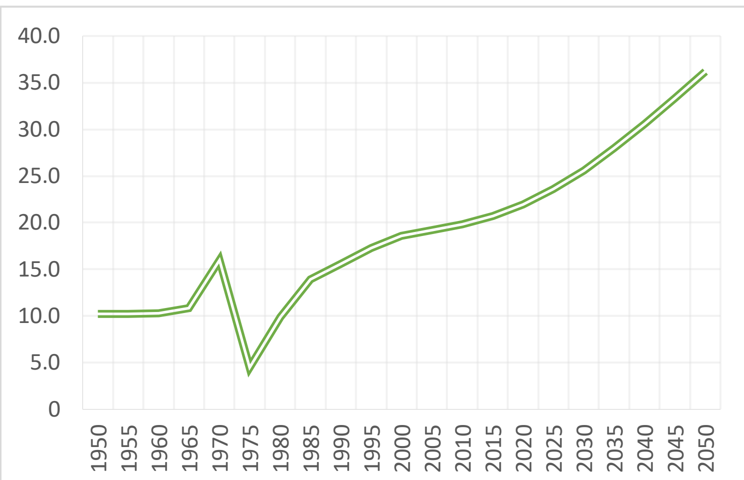
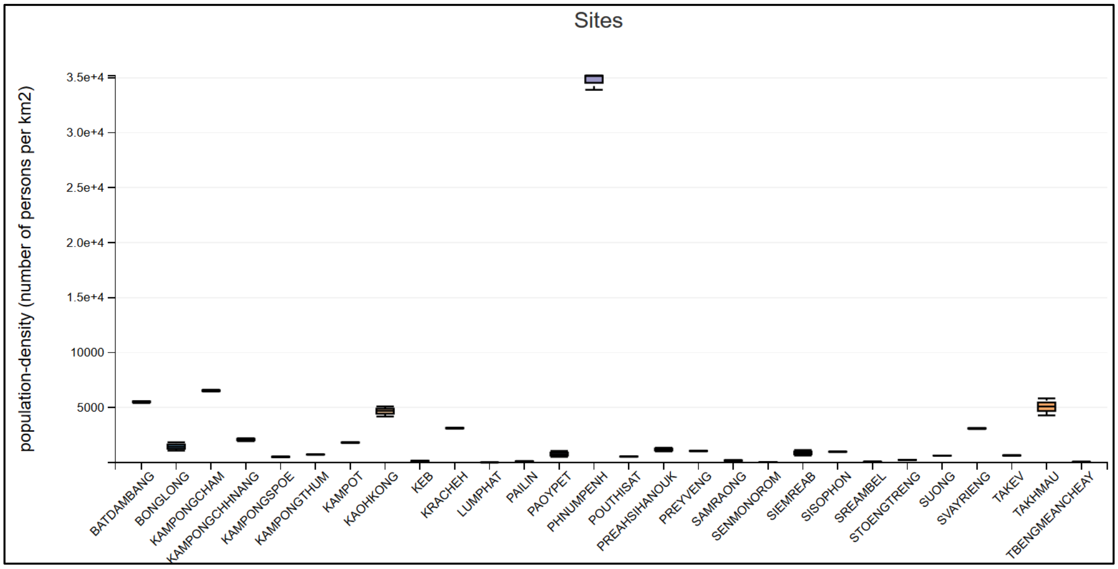


Figure 6: Proportion urban in Cambodia (WUP 2014)

This example first considers the whole country's urban system. Point data sample is extracted using the file upload option for Cambodian settlement points (from SEDAC's Global Rural-Urban Mapping Project or GRUMP) and selecting AppEEARS layers for population counts and density (GPW4) and vegetation cover (MODIS).

Figures 7a and 7b below show one of the data visualization options for these two selected layers. All the Cambodian settlement points are plotted using boxplots that summarize the data points available in each layer (3 for population counts, and 10 for vegetation cover). Note that the population density is not the average density of the whole city, but the density at the city centroid (settlement point) where the percent vegetation is also measured. In the figure, the capital, Phnom Pen, has the highest population density and the one lowest percentage of tree cover area.



b)

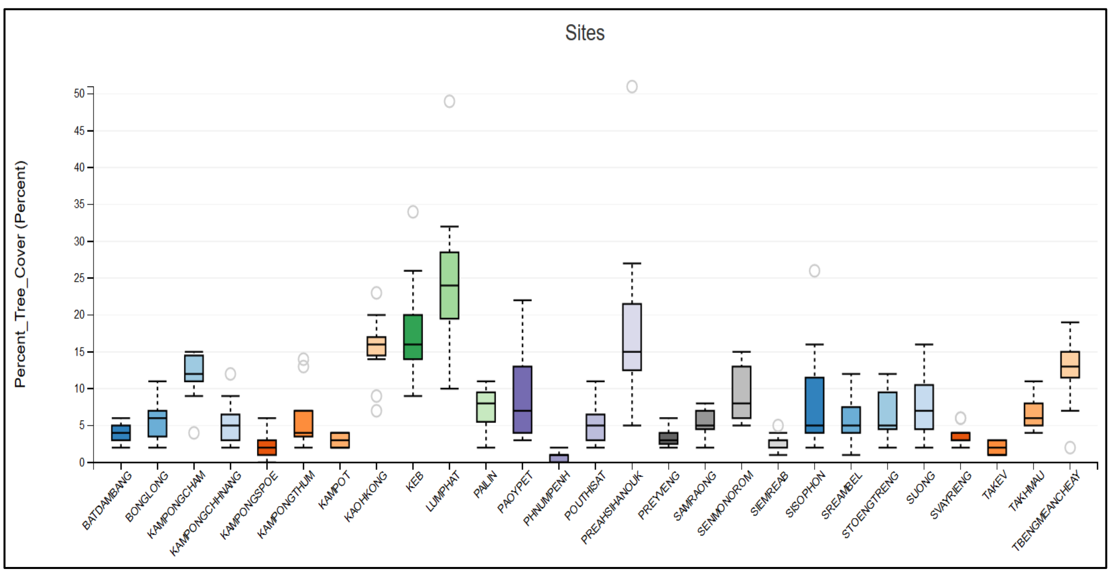


Figure 7: Cambodia's urban system depicted using settlement points from GRUMP: (a) total population 2000, 2005 and 2010 (from GPW4), and (b) percent of tree cover 2000-2010

The output dataset (in csv format) attaches covariates (in green) to the point data sample used for the data extraction (in orange), as show in Table 1 for two of Cambodian settlement points. Again, the counts and density are only representative of the approximately one square kilometer grid cell where the settlement point is located.

Table 1: Selected urban settlements in Cambodia with total population 2000, 2005 and 2010 (from GPW4), and (b) percent of tree cover 2000-2010

Settlement point ID	Latitude	Longitude	Year	Population count	Population density (pop/sq km)	Non tree vegetation (%)	Non Vegetated (%)	Tree cover (%)
BATDAMBANG	13.1	103.2	2000	4533	5408.68	49	45	6
BATDAMBANG	13.1	103.2	2005	4675	5577.54	60	34	6
BATDAMBANG	13.1	103.2	2010	4647	5544.67	34	63	3
BONGLONG	13.74165	106.9969	2000	903	1079.99	83	11	6
BONGLONG	13.74165	106.9969	2005	1199	1434.71	84	10	6
BONGLONG	13.74165	106.9969	2010	1536	1837.39	75	14	11

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. However, for reasons alluded to above, it may be wiser to use the area representation of urban settlements (as shown in example 2).

2. Changes in population density and vegetation cover in Cambodian urban extents
This example moves from points to areas, uploading a shapefile containing Cambodian urban extents (from GRUMP) for the same covariates (population and vegetation cover).

The outcome data for the urban extent containing the capital Phnom Penh and the city of Takhmau are presented below (Figure 8). It seems to be a direct relationship between changes in population density and changes in the percent not vegetated.

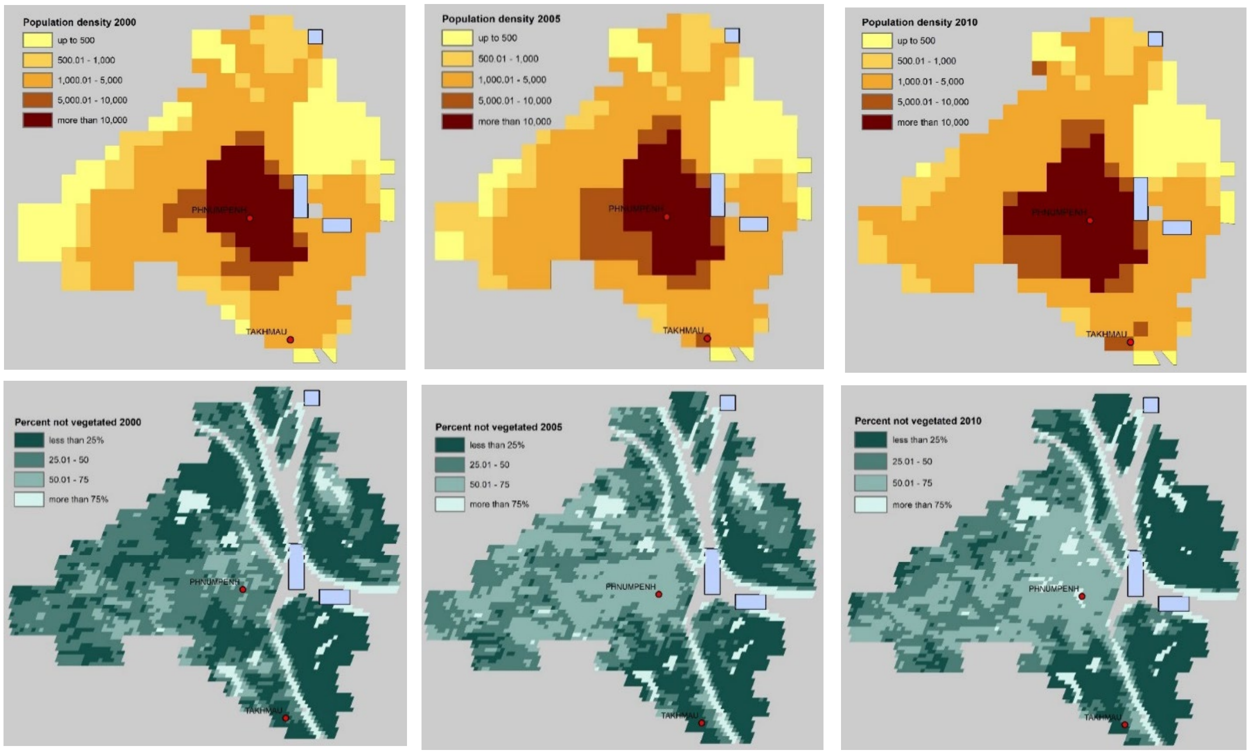


Figure 8: Phnom Penh and Takhmau: changes in population density (upper panel) and vegetation cover (MODIS) (lower panel), 2000, 2005 and 2010

Acknowledgments

This webpage is partially based on background briefs to the Population-Environment Research Network (PERN) Cyberseminar on "People and Pixels Revisited: 20 years of progress and new tools for population-environment research" by A. de Sherbinin and S. Adamo (GIOVANNI and AppEEARS) and T. Kugler (IPUMS Terra). De Sherbinin and Adamo are grateful to D. Meyer and M. 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 to their brief.

Selected references on Giovanni

Acker, J., Soebiyanto, R., Kiang, R., & Kempler, S. (2014). Use of the NASA Giovanni Data System for Geospatial Public Health Research: Example of Weather-Influenza Connection. *ISPRS International Journal of Geo-Information, 3*(4), 1372.

Chudnovsky, A. A., Koutrakis, P., Kostinski, A., Proctor, S. P., & Garshick, E. (2017). Spatial and temporal variability in desert dust and anthropogenic pollution in Iraq, 1997–2010. *Journal of the Air & Waste Management Association, 67*(1), 17-26. doi:10.1080/10962247.2016.1153528

Eisenberg, M. C., Kujbida, G., Tuite, A. R., Fisman, D. N., & Tien, J. H. (2013). Examining rainfall and cholera dynamics in Haiti using statistical and dynamic modeling approaches. *Epidemics, 5*(4), 197-207. doi:https://doi.org/10.1016/j.epidem.2013.09.004

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.

Prados, A. I., Leptoukh, G., Lynnes, C., Johnson, J., Rui, H., Chen, A., & Husar, R. B. (2010). Access, Visualization, and Interoperability of Air Quality Remote Sensing Data Sets via the Giovanni Online Tool. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3*(3), 359-370. doi:10.1109/JSTARS.2010.2047940

Selected references on AppEEARS

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

Cited 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