Time Series Analisys on Geoespatial Data with Python

Author: João Otavio Nascimento Firigato

email: joaootavionf007@gmail.com

LinkedIn: https://www.linkedin.com/in/jo%C3%A3o-otavio-firigato-4876b3aa/

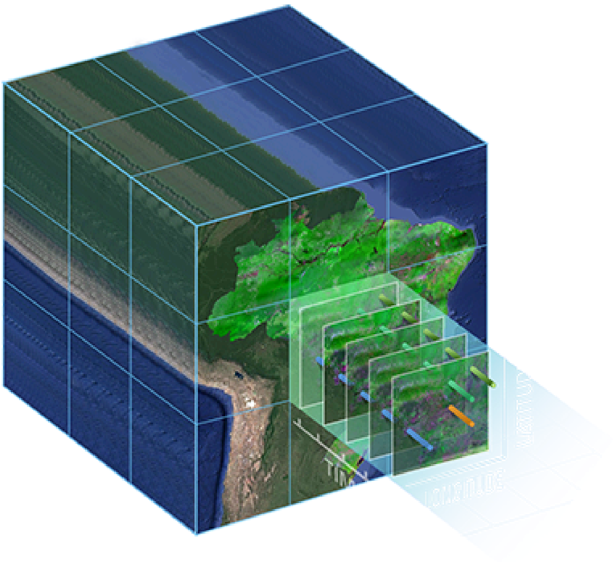
First instructions:

✅ Access the link to join our private WhatsApp community for students: https://chat.whatsapp.com/EPn27ZgR07lF3e1vnj8FiI

❗ It is important to access the Whatsapp Group to get the Colab Notebooks, as the PDF files are protected from text copying.

Chapter 2 - Time Series with Geospatial Data

Time series analysis of geospatial data allows us to analyze and understand how events and attributes of a place change over time. Its use cases are broad, particularly in social, demographic, environmental, and meteorological/climatic studies. In environmental sciences, for example, time series analysis helps analyze how the land cover/land use of an area changes over time and its underlying drivers. It is also useful in meteorological studies to understand spatiotemporal changes in weather patterns (I will soon demonstrate one such case study using precipitation data). Social and economic sciences benefit greatly from this analysis to understand the dynamics of temporal and spatial phenomena, such as demographic, economic, and political patterns.



Important considerations

One of the most persistent problems in geospatial analysis is consistency in time-series analysis. We can't help but want to compare the growth of one city to another over time. Time is the easy part—a year today is the same as a year millennia ago—but spatial definitions? A whole different question.

A few weeks ago, we looked at manufacturing employment patterns, and a natural follow-up is to ask: In areas where manufacturing jobs have disappeared, have residents' incomes declined? Or have these traditionally lower-middle-income jobs been replaced by higher-income sectors?

This requires a definition of "place" that is appropriate to our topic of study—one that needs to consider several factors:

How many geographic areas are there?

Too few, and the study becomes uninteresting. Too many, and the forest may be lost for the trees – local detail swamps the broader geographic patterns we want to explore.

How much interaction is there between the areas?

Even at the state level, we face the annoying problem that people insist on working in a state other than the one they live in. While some level of interaction between fields of study is always present, choose fields that are too small and the interaction effects will dominate the analysis.

How stable are area definitions over time?

The areal extent of a spatial object can change, even if it continues to have the same name. Postal codes and city boundaries are notorious for changing over time. While instability is not an insurmountable problem, it is inevitable that the most interesting areas where change is occurring will be the most unstable.

The analyst must be aware of territorial changes, especially for areas with consistent naming.

Is data collected or disseminated at this scale?

Any re-estimation of spatial data boundaries will result in increased error, and this should be avoided where possible or minimized. As a general rule, data should be available for more detailed spatial areas than those analyzed in order to minimize the effects of changing definitions over time.

Do these areas cover the entire country?

The most recognizable areas are cities and metropolitan areas, but these do not extend beyond urbanized areas or are defined using administrative boundaries, such as counties, that bear little resemblance to the areas they are named after.

How recognizable are the areas?

We could easily define arbitrary areas as grids or hexagonal shapes, but readers of the analysis will need to assign meaning to these grids by cluttering the maps with labels. By ignoring natural boundaries (coastlines, lakes), these areas create interpretation problems when cropped.

Time Series in Remote Sensing

Time series analysis of satellite imagery is a valuable tool for investigating the dynamic processes at work on our planet. We can gain insights into many different fields, from agriculture and forestry to urban planning, by collecting and analyzing images captured over time.

A satellite imagery time series is a sequence of images (data points) of the same location recorded at regular intervals over a given study period. Here, time is a critical factor that adds depth to the available satellite data because it reveals not only the final results but also the changes during the interval of the data points.

Time series analysis is a common use of the collected information. It typically requires many satellite observations to ensure the validity, accuracy, and consistency of the data. Working with large data sets provides the option to eliminate random effects and correct for seasonal fluctuations. Meanwhile, detecting trends and seasonal changes in satellite imagery time series provides additional useful information for making more informed decisions. Real-world applications of time series satellite remote sensing data include using past events as a basis for making predictions about the future.

Let's say that at the beginning of the growing season, you plan to build a time series graph based on historical satellite imagery. It's a shock to learn that one of your fields is producing significantly less than it did in previous years. The key to avoiding future devastating crop losses is to identify the problem as quickly as possible so that its source can be isolated and eradicated. Long-term time series analysis of satellite imagery can sort through tons of data collected over several years, apply the appropriate vegetation indices to each image, and help you identify trends in this mountain of data.

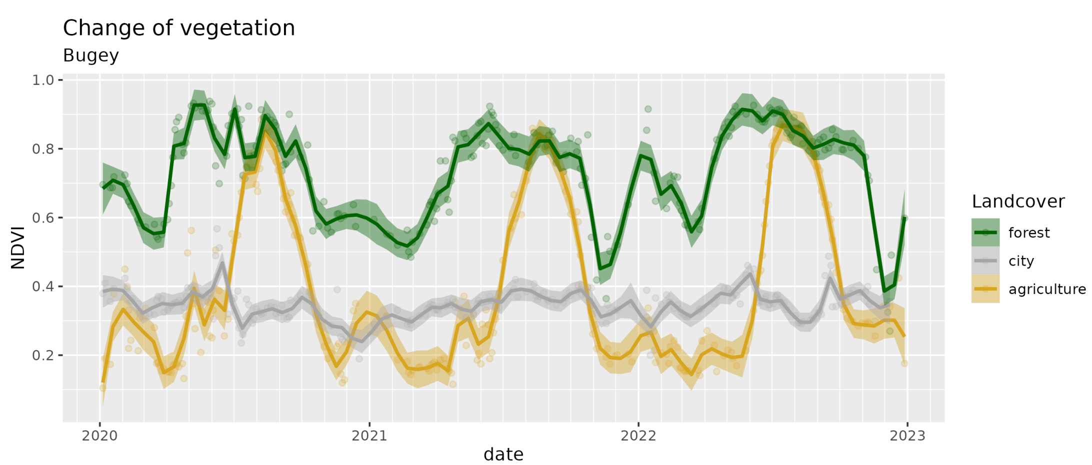
By continuously accumulating and monitoring satellite imagery, you can quickly and accurately assess changes, recognize trends and patterns, spot anomalies, and make sound business decisions and corrective actions in your areas of interest (AOI).

Application areas for satellite data time series

Time series of satellite imagery provide invaluable information for countless applications. The common thread among these uses is the notion that satellite data can help us learn more about the ways in which the planet is changing, the forces behind these changes, and their potential repercussions.

Agriculture

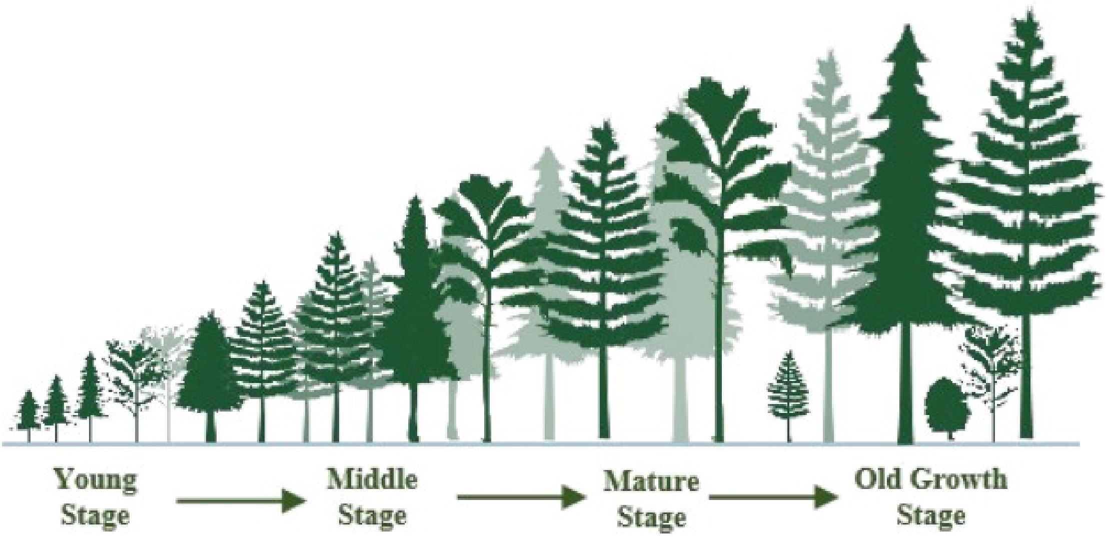
Farmers can benefit greatly from using satellite imagery time series as a source of valuable information about the growth and yield dynamics of their crops, as well as an aid in forecasting crop production and agricultural market prices. To avoid economic losses caused by insect infestations, satellite imagery time series are used for pest and crop health monitoring. By knowing the vegetation history of their crops, farmers can optimize the allocation of inputs (water, nutrients, chemicals, seeds, and more) and improve the planning of their farm management activities. Crop monitoring, crop rotation planning, irrigation and fertilizer control, pest and disease management, yield forecasting, and land use management are just some of the many agricultural uses for satellite data time series analysis.



Forestry

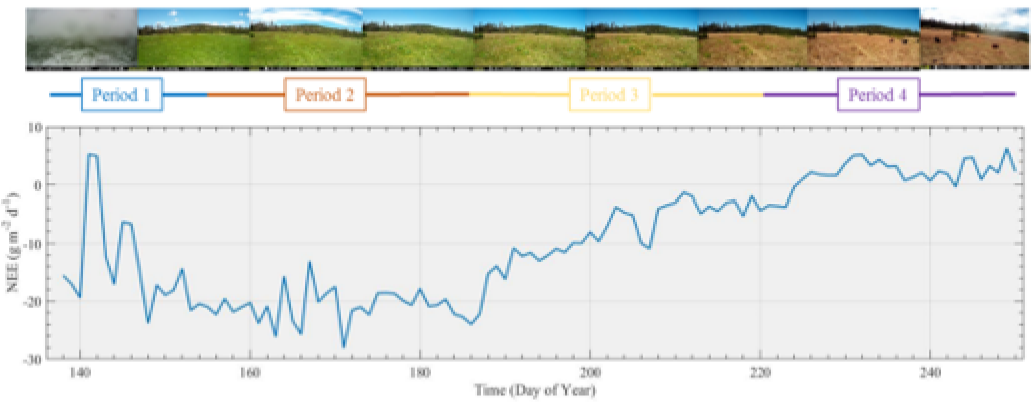
The public release of the Landsat archive, which is known as the most widely used dataset for time-series analysis of satellite imagery, has lowered the barrier to entry for forest monitoring. The past decade has also seen further advances in forest change detection technology, enabling more in-depth monitoring of forests.

Satellite data series can help locate problem areas and inform decision-making by observing changes in the state and area of forest cover, overall forest health and composition over time. In particular, by analyzing satellite imagery time series, forest managers can identify deforestation hotspots and measure the rate at which trees are being cut. Analysis of satellite data series can therefore inform afforestation, reforestation, forest restoration and other forest management strategies.



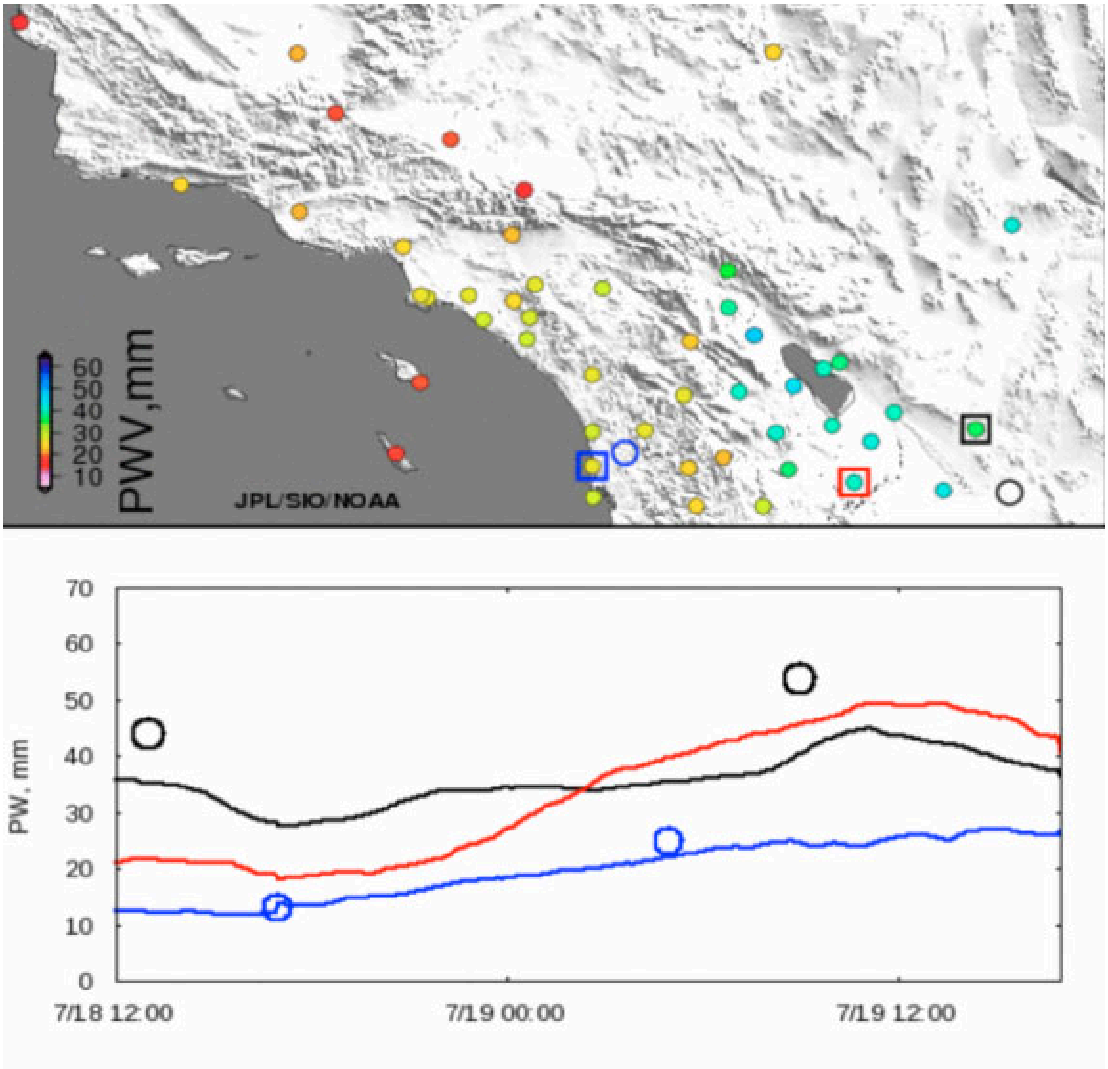
Climate Change Monitoring

Time series of satellite images, the backbone of statistical climate analysis, can produce actionable insights for climate study and policymaking. Computer models of climate systems based on time series of satellite images can be used to monitor Earth's radiation budget, examine the effects of climate change on plant life, monitor sea level, observe the movement of glaciers and ice caps, and much more. For example, the GEOS-5 Atmosphere-Ocean General Circulation Model (AOGCM) is a state-of-the-art model that seeks to reflect climate variability over many time periods (from synoptic to multisecular).



Disaster Management

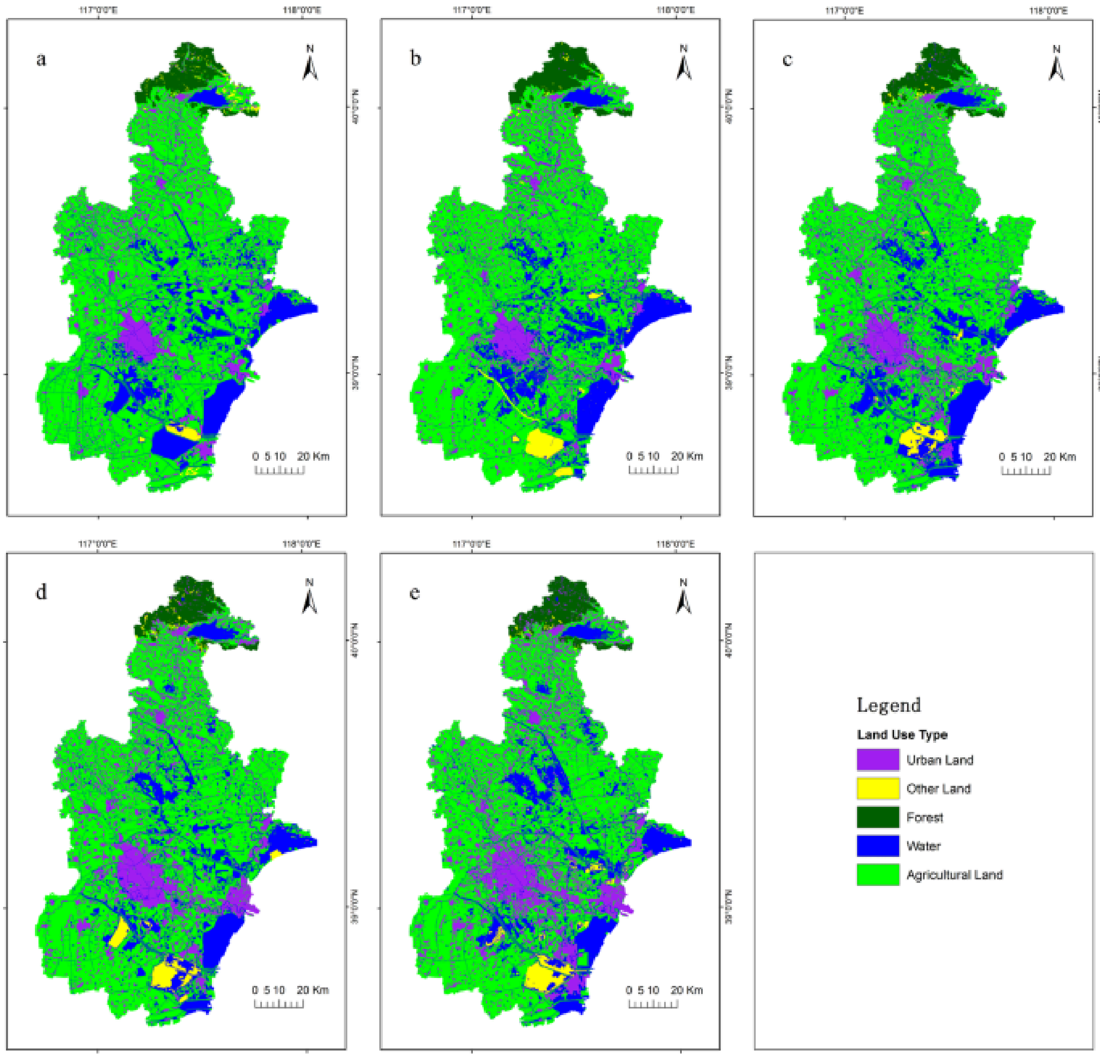
Time series analysis has many applications in emergency management, including heads-up detection and disaster risk prediction. Satellite imagery time series also help assess the devastation caused by natural disasters such as earthquakes, storms, floods, landslides, and wildfires (more in our 2023 natural disasters article). With the help of this data, disaster management managers can better coordinate response efforts and allocate available resources more efficiently, as well as stay up-to-date on the progress of restorations in damaged areas.



Land use planning

Time series analysis of satellite images can shed light on land use change and the spread of cities. This knowledge is essential in the context of sustainable land use, where the appropriate configuration of multiple uses (agriculture, forestry, cities, etc.) ensures long-term soil conservation and ecosystem balance. Planning and management using time series of satellite images can improve urban construction, infrastructure, and municipal services. In addition, satellite images help track and analyze the effectiveness of urban transportation systems.

This is just a small sample of the many real-world applications of time series of satellite images. Furthermore, the use of satellite data to gain insights and act on global concerns will grow exponentially as technology advances.



Limitations in remote sensing time series

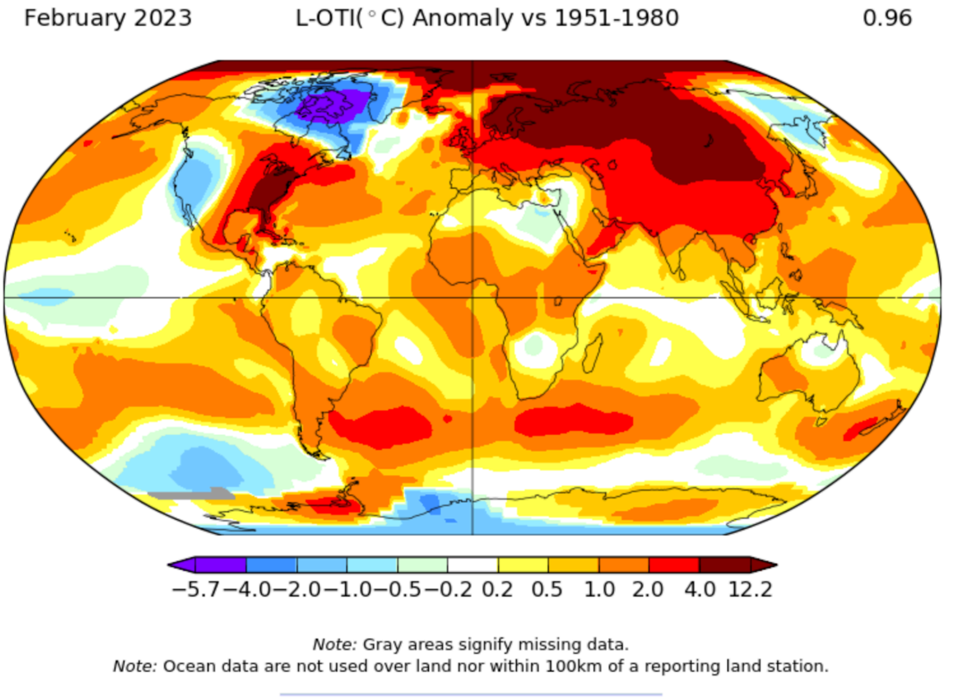
Remote sensing time series face several limitations that make modeling these data challenging. One of the main challenges is the lack of consistency and accuracy in the data collected. Unlike conventional time series, where data are regularly measured and generally complete, remote sensing data are subject to errors and gaps, such as the removal of cloudy pixels, which can result in the loss of a significant portion of the data. In addition, technical problems, such as instrument malfunctions, can compromise the cohesion of time series, especially over long observation periods. These factors make it difficult to build explanatory models that are both simple and capable of capturing all the variations in the data.

Another important limitation is the variation in the scale and magnitude of measurements over time. In remote sensing time series, it is common for measured values to be influenced by external conditions, such as snowfall or atmospheric distortions, which can significantly alter the interpretation of the data. For example, changes in atmospheric conditions can affect the visual appearance of images, distorting the color or brightness of vegetation. This raises questions about the comparability of data over time and whether certain values should be retained or discarded in the analysis. These variables make the process of analyzing and interpreting remote sensing time series more complex and prone to error, requiring more sophisticated approaches to deal with these inconsistencies.

Time series with climate data

Time series of climate data are essential for studying and understanding atmospheric variations and trends over time. These data, collected at regular intervals, allow us to analyze seasonal patterns, identify anomalies, and predict future climate change. For example, time series analysis can reveal global warming trends, variations in precipitation, and changes in wind patterns, providing a detailed view of the effects of climate change in different regions of the planet.

The use of climate time series is essential for modeling and simulating future scenarios, helping to formulate environmental policies and mitigation strategies. In addition, these data are crucial for sectors such as agriculture, energy, and water resource management, which depend on climate forecasts for planning and operations. Accurate analysis of these time series allows for better adaptation to climate change, contributing to community resilience and environmental preservation.



The main types of climate data time series include:

- Temperature: Temperature time series record daily, monthly or annual variations in air temperature in a given location. These series are essential for studying warming or cooling trends, identifying heat and cold waves, and understanding seasonal variations.
- Precipitation: These series record the amount of rain, snow or hail that falls in a region over time. They are essential for analyzing patterns of drought, flooding and

changes in precipitation distribution, directly influencing agriculture and water supply.

- Humidity: Humidity time series measure the water vapor content in the atmosphere. Relative humidity is often monitored, as it has a significant impact on climate, cloud formation and the perception of heat.
- Atmospheric pressure: These series record the pressure exerted by the atmosphere on a given area over time. Atmospheric pressure is crucial for predicting changes in the weather, such as the arrival of cold fronts and the formation of storms.
- Wind speed and direction: Wind time series capture variations in wind speed and direction, which are important for understanding atmospheric circulation patterns, influences on regional climate, and for modeling pollutant dispersion.
- Solar radiation: This type of series measures the amount of solar energy received on a surface over time. It is essential for the study of energy balances, local climate, and for planning solar energy systems.

These types of time series, when analyzed together, provide a comprehensive view of climate conditions and their changes over time, allowing a deeper understanding of the interactions between different elements of the climate system.

Main Time Series Databases

Google Earth Engine is a cloud-based geospatial analytics platform that allows users to visualize and analyze satellite imagery of our planet. GEE is the leading cloud-based geospatial data processing and analytics platform available today, and it is completely free to use for students, researchers, and developers. While Javascript is the most widely used language in GEE, fortunately there is also the option to use the API with Python.

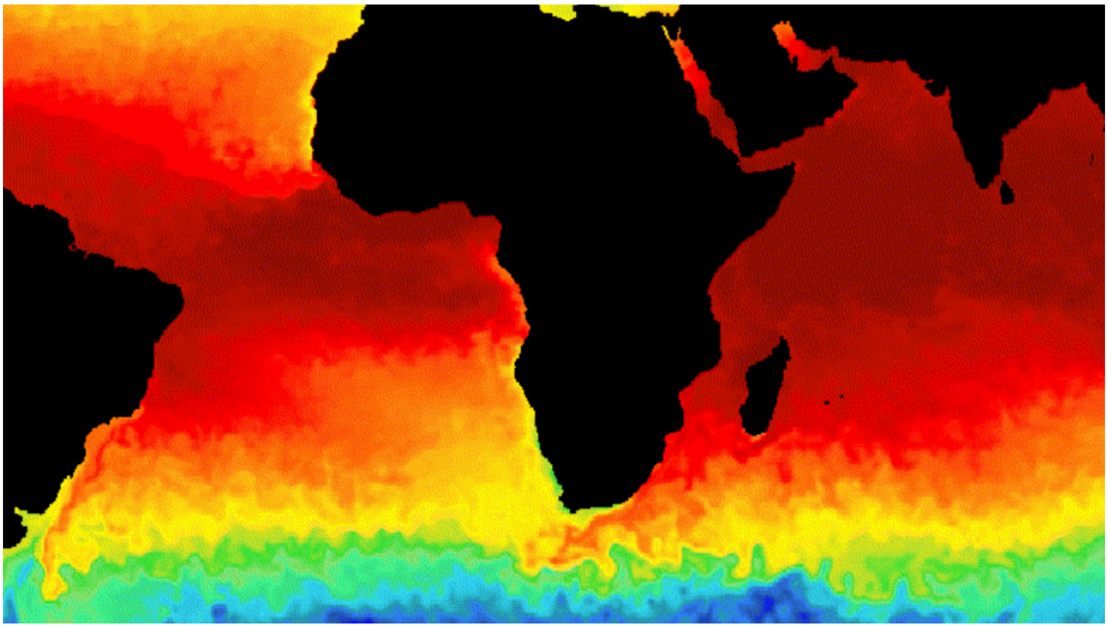
GEE Data Catalog

GEE provides direct access to a multi-petabyte catalog of satellite imagery and geospatial datasets, including the entire Landsat catalog from EROS (USGS/NASA), MODIS, and

Sentinel-2 imagery, as well as precipitation, elevation, sea surface temperature, and climate data.

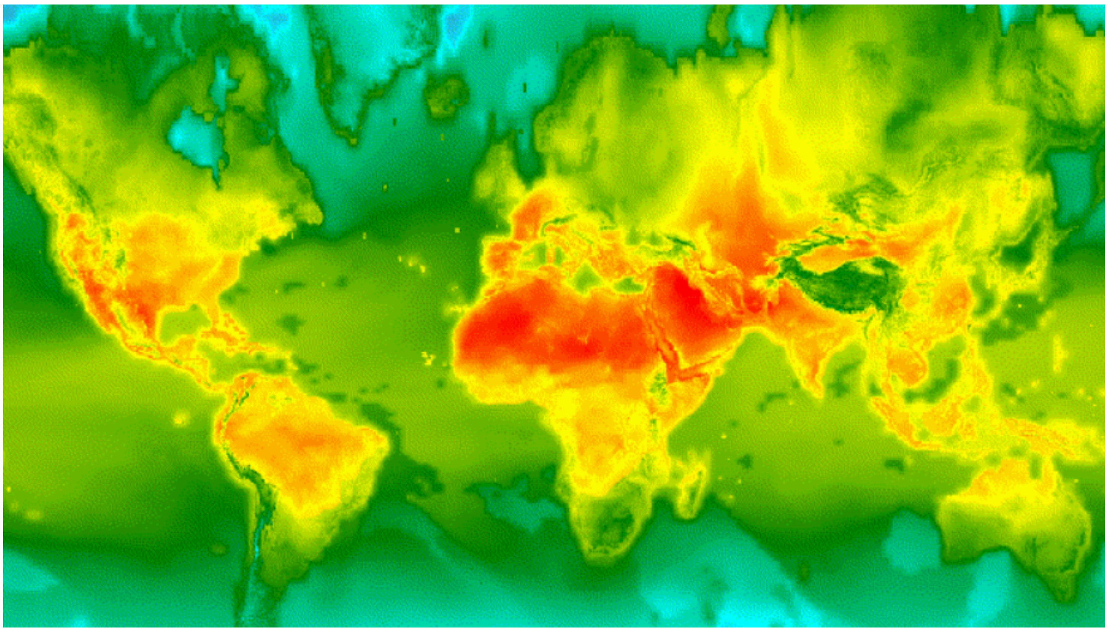
Surface Temperature

Thermal satellite sensors can provide surface temperature and emissivity information. The Earth Engine data catalog includes land and sea surface temperature products derived from several spacecraft sensors, including MODIS, ASTER, and AVHRR, as well as raw thermal data from Landsat.



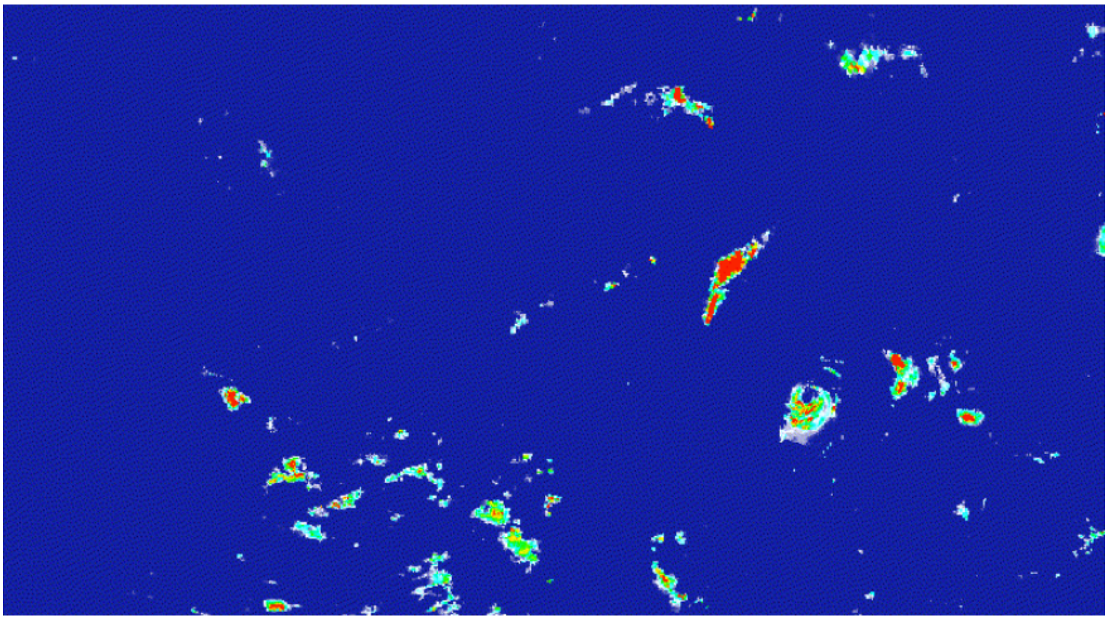
Climate

Climate models generate long-term climate forecasts and historical interpolations of surface variables. The Earth Engine catalog includes historical reanalysis data from NCEP/NCAR, gridded meteorological datasets such as NLDAS-2 and GridMET, and output from climate models such as the University of Idaho's MACAv2-METDATA and NASA Earth Exchange's downscaled climate projections.



Meteorology

Weather datasets describe predicted and measured conditions over short periods of time, including precipitation, temperature, humidity, wind, and other variables. Earth Engine includes forecast data from NOAA's Global Forecast System (GFS) and NCEP Climate Forecast System (CFSv2), as well as sensor data from sources such as the Tropical Rainfall Measuring Mission (TRMM).

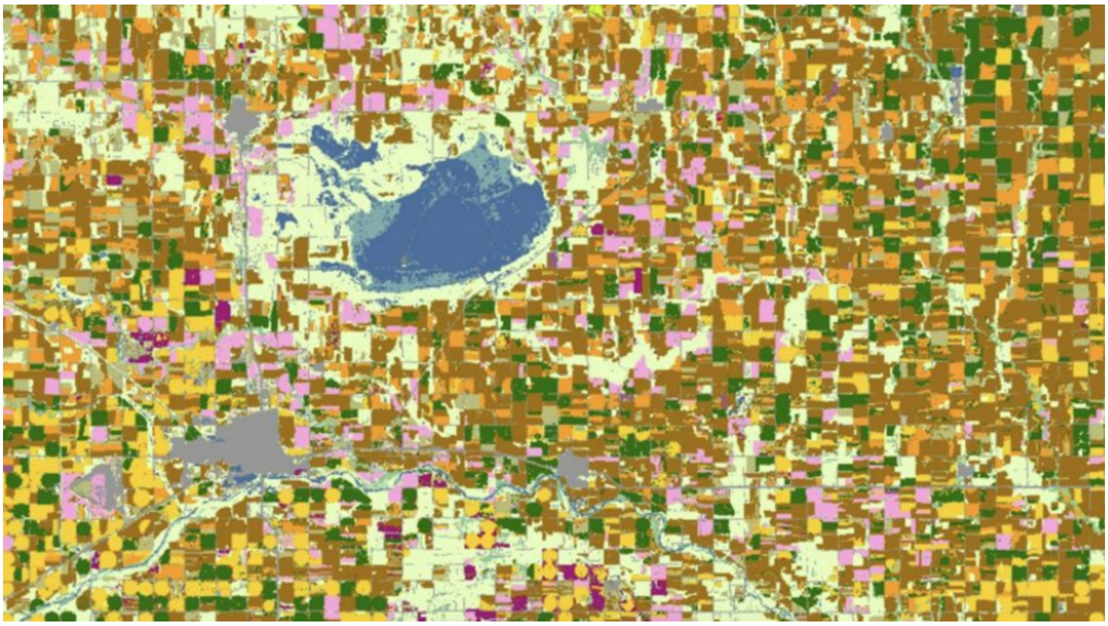


Optical

Landsat, a joint program of the USGS and NASA, has been observing Earth continuously since 1972 to the present day. Today, Landsat satellites image the entire Earth's surface at 30-meter resolution approximately once every two weeks, including multispectral and thermal data. The Copernicus Program is an ambitious initiative led by the European Commission in partnership with the European Space Agency (ESA). The Sentinels include all-weather radar imagery from Sentinel-1A and -1B, high-resolution optical imagery from Sentinel 2A and 2B, and ocean and land data suitable for environmental and climate monitoring from Sentinel 3. The Moderate Resolution Imaging Spectroradiometer (MODIS) sensors on NASA's Terra and Aqua satellites have been acquiring daily Earth imagery since 1999, including daily imagery, 16-day BRDF-adjusted surface reflectance, and derived products such as vegetation and snow cover indices.

Agricultural Data

Cropland data is essential to understanding global water consumption and agricultural production. Earth Engine includes several cropland data products, such as the USDA NASS Cropland Data Layers, as well as layers from the Global Food Security-Support Analysis Data (GFSAD), including cropland extent, crop dominance, and water sources.



Thank you! See you in the next Chapter!