Giovanni:
Examining NASA Remote-Sensing Data for Public Health

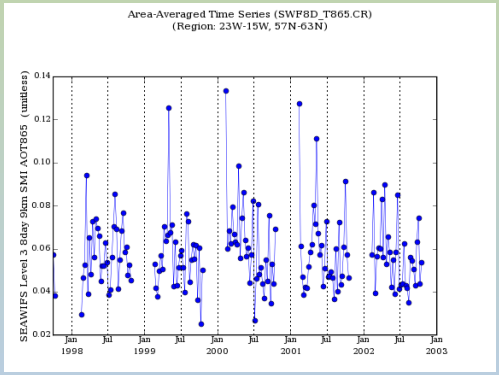
James G. Acker, NASA GES DISC / Adnet Inc.
Radina Soebiyanto, USRA
August 27, 2013

2nd Symposium on Advances in Geospatial Technologies for Health
MEDGEO 2013

Giovanni is…

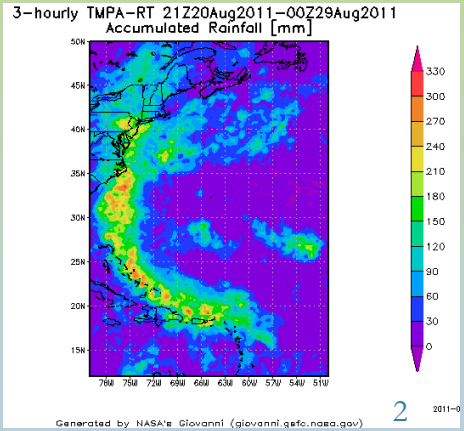
… The **G**eospatial **I**nteractive **O**nline **V**isualization **AN**d a**N**alysis **I**nterface.

Since 2003, Giovanni has provided access to a wide variety of NASA remote sensing data and related data sets, allowing many different kinds of researchers to use NASA data.



Time-series of aerosol optical thickness south of Iceland

Accumulated rainfall from Hurricane Irene



Part 1:
The Powers of Giovanni

Giovanni has been used widely for scientific research for several reasons:

- Ease of access to many different kinds of remotely-sensed and model data products
- No need for additional software or tools to read and plot the data
- Rapid generation of data plots
- Immediate download of results, both as data files and plots
- Many different kinds of data visualizations

Getting Started with Giovanni – the current data portal interface

Select Area of Interest

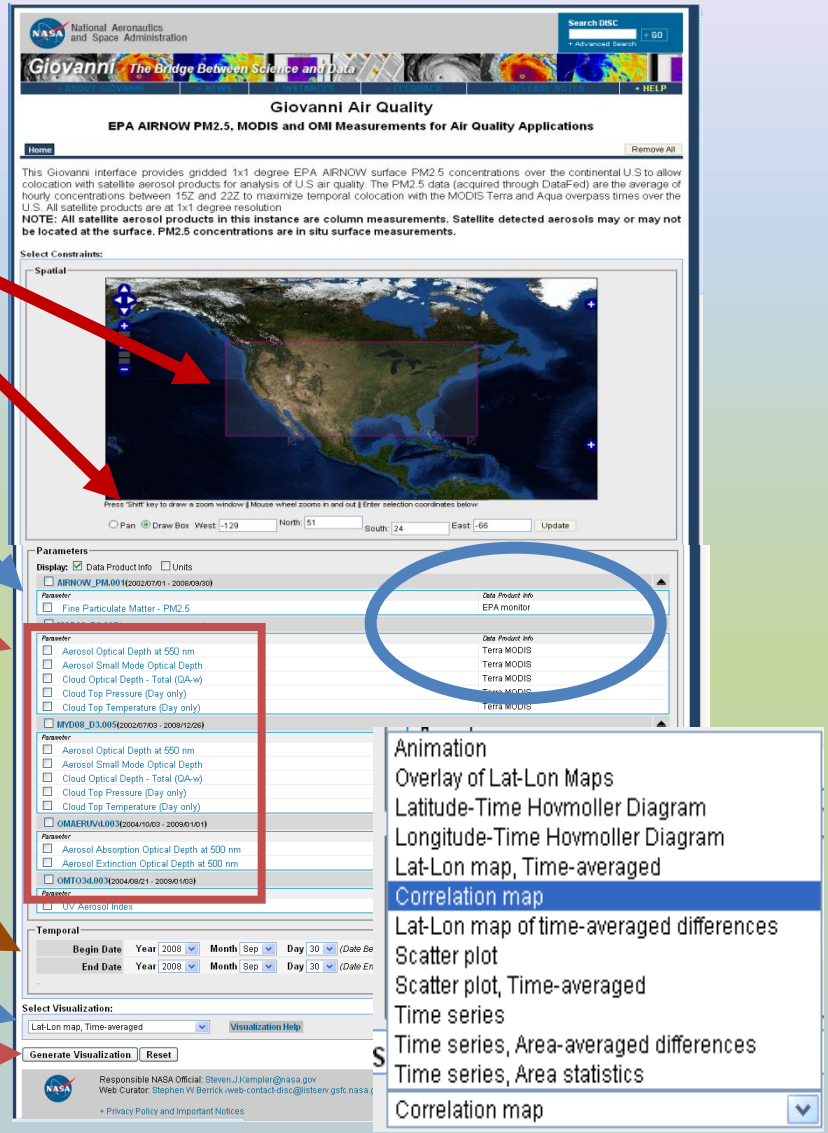
Select Display (info, unit)

Select Parameters

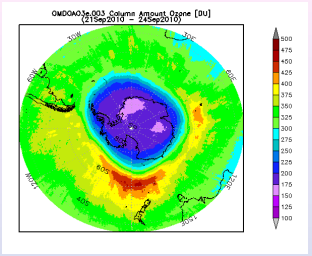
Select Time Period

Select Plot type

Generate Visualization



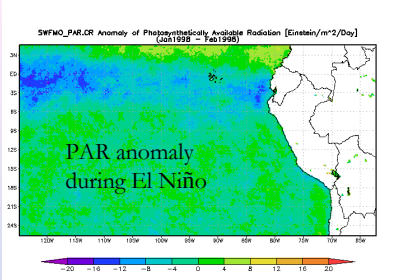
The power of visualization



While Giovanni has been used for many different kinds of research, it was primarily envisioned to be a data *exploration* tool. The main data it serves are Level 3 data products, which are lower spatial resolution gridded global data.

Giovanni allows users to make and 'tweak' maps and plots rapidly, indicating potentially fruitful research areas. Research may then be conducted with higher spatial resolution data.

Thus, Giovanni's variety of visualizations is one of its main analytical powers.



The current Giovanni visualization suite

In the following slides, the suite of visualization options available in Giovanni will be shown. They include:

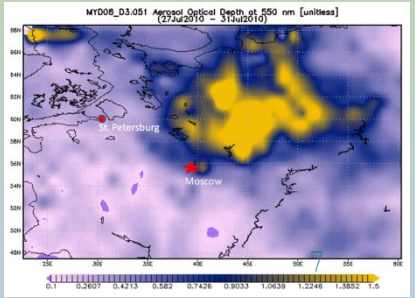
Data maps

- Lat-lon maps, time-averaged
- Correlation maps
- Difference maps
- Anomaly maps
- Animations
- KMZ file option

Data plots

- Time-series
- Hovmöller diagrams
- X-Y scatterplots
- Vertical profiles
- Histograms

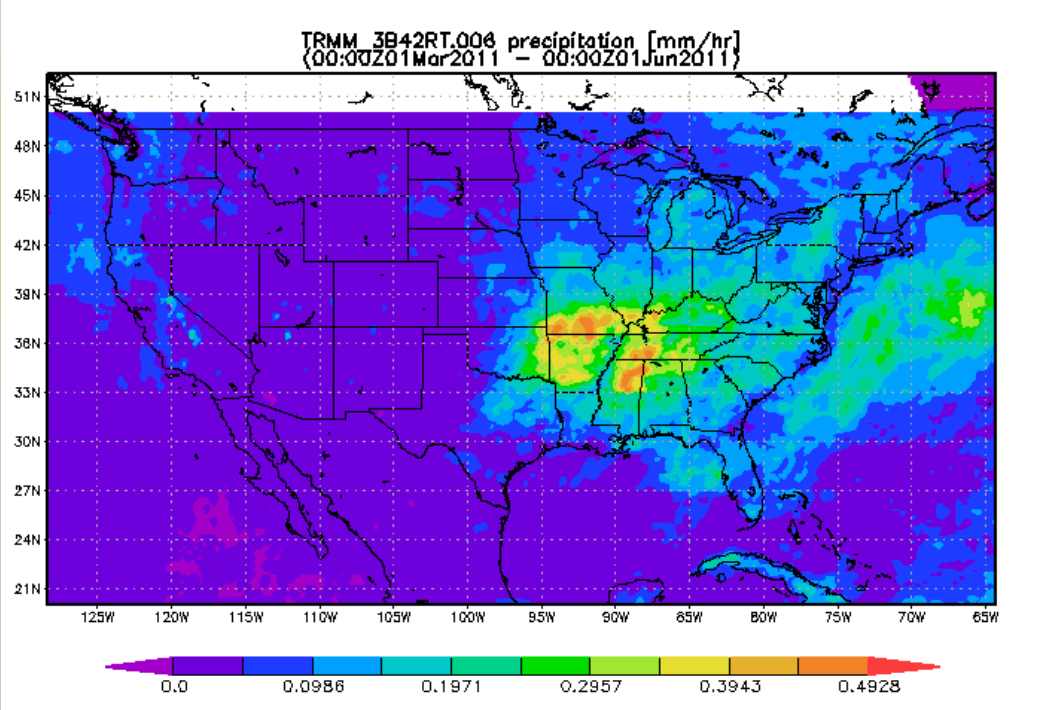
Wildfire smoke over Russia



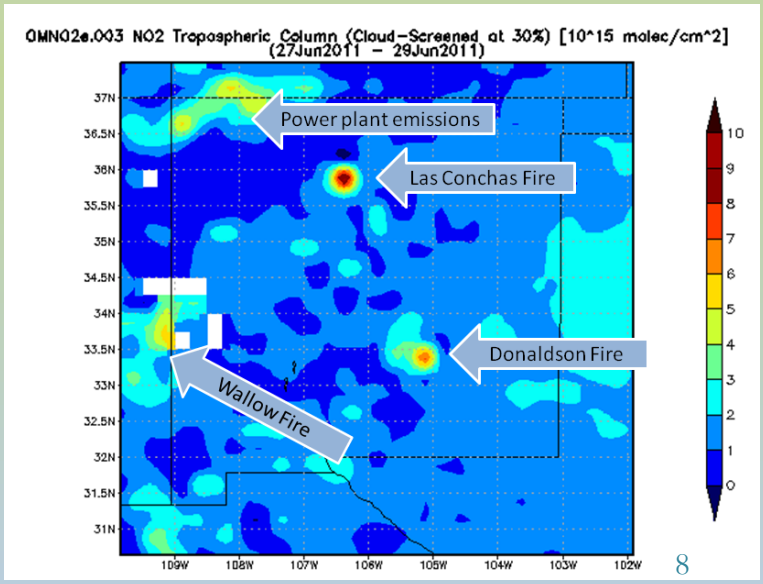
Lat-Lon maps, time-averaged

Giovanni's most basic visualization is the ***data map:*** data values represented on a global or regional map, represented with a false color palette. Data can be shown for a single time increment, or averaged over a time range.

TRMM precipitation, March-June 2011



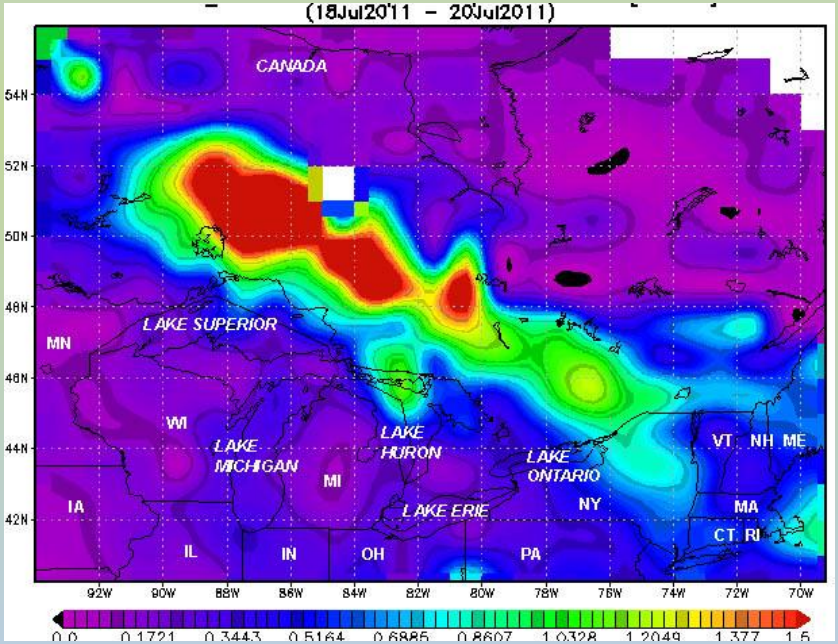
OMI NO_2, June 27-29, 2011 annotated in Powerpoint



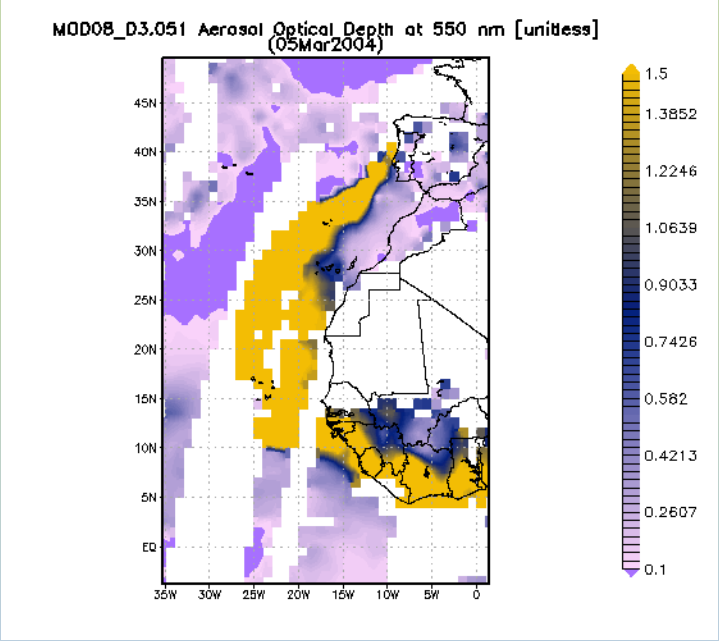
Easy tweaks: color palettes, data ranges

Giovanni allows users to change color palettes, or change the maximum and minimum values of the color palette range, to emphasize features in the data.

MODIS Aerosol Optical Depth, June 18-20, 2011, Purple-Red + Stripes palette



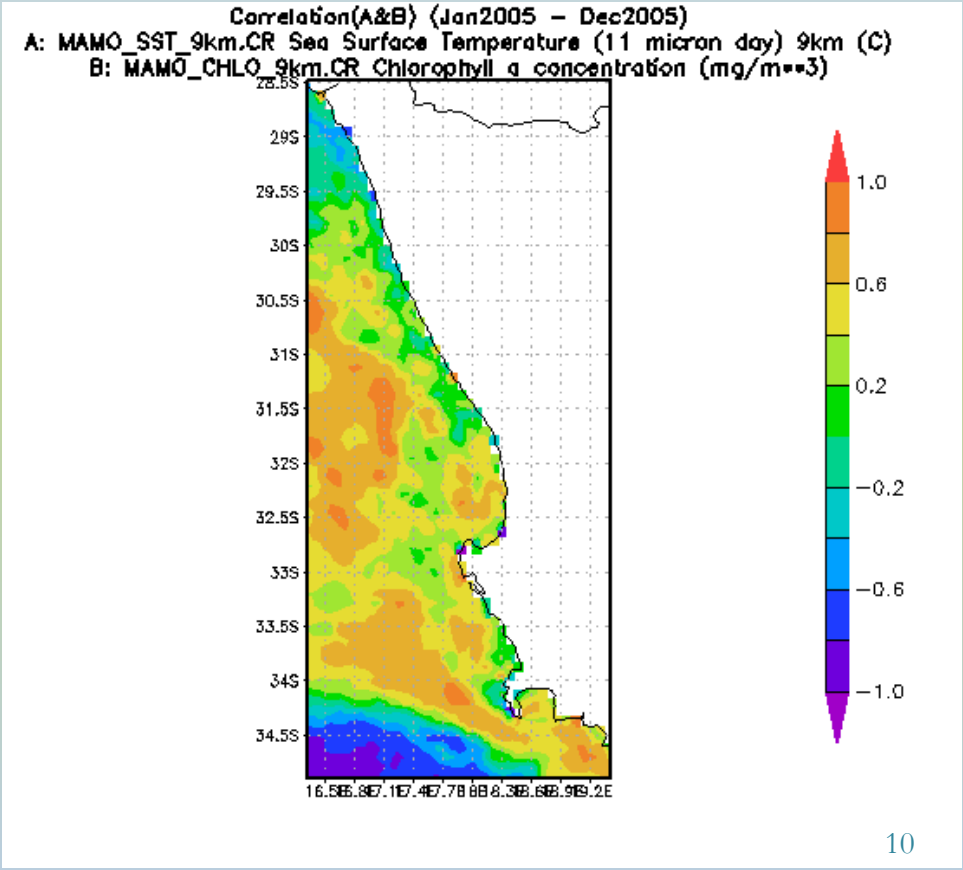
MODIS Aerosol Optical Depth, March 5, 2004, Haze palette, custom range



Correlation maps

Correlation maps show where data are correlated over time, i.e. where similar data values for different data products occur together. Such maps can be used for examining potential cause-and-effect relationships.

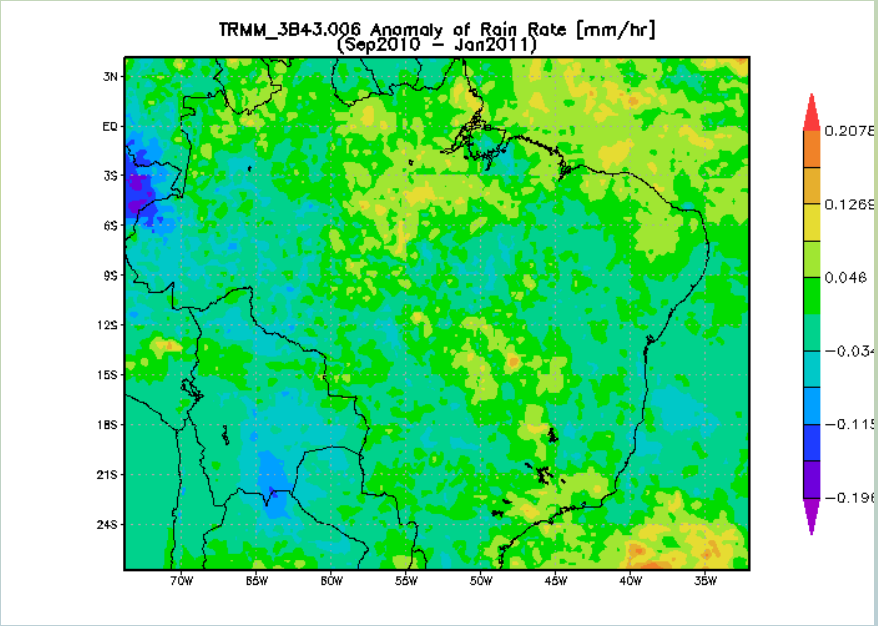
Correlation map of MODIS sea surface temperature and chlorophyll *a* concentration in the Benguela Upwelling Zone off the southwest coast of Africa. Where chlorophyll and SST do not vary much (offshore), the correlation is high. In the transition area, because chlorophyll and SST are more variable, the correlation is lower. Where upwelling is occurring, SST and chlorophyll will be negatively correlated. This map is for the year 2005.



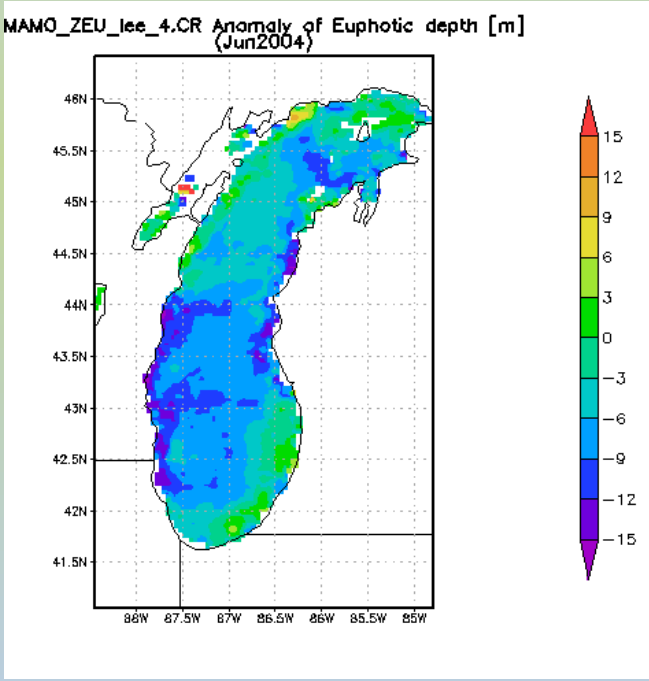
Anomaly maps

Provided that a climatology is available, anomaly maps are an excellent way to display events that are departures from 'normal' climate and environmental conditions. Climatologies are created by the data providers.

TRMM precipitation anomaly for South America, September 2010 – January 2011



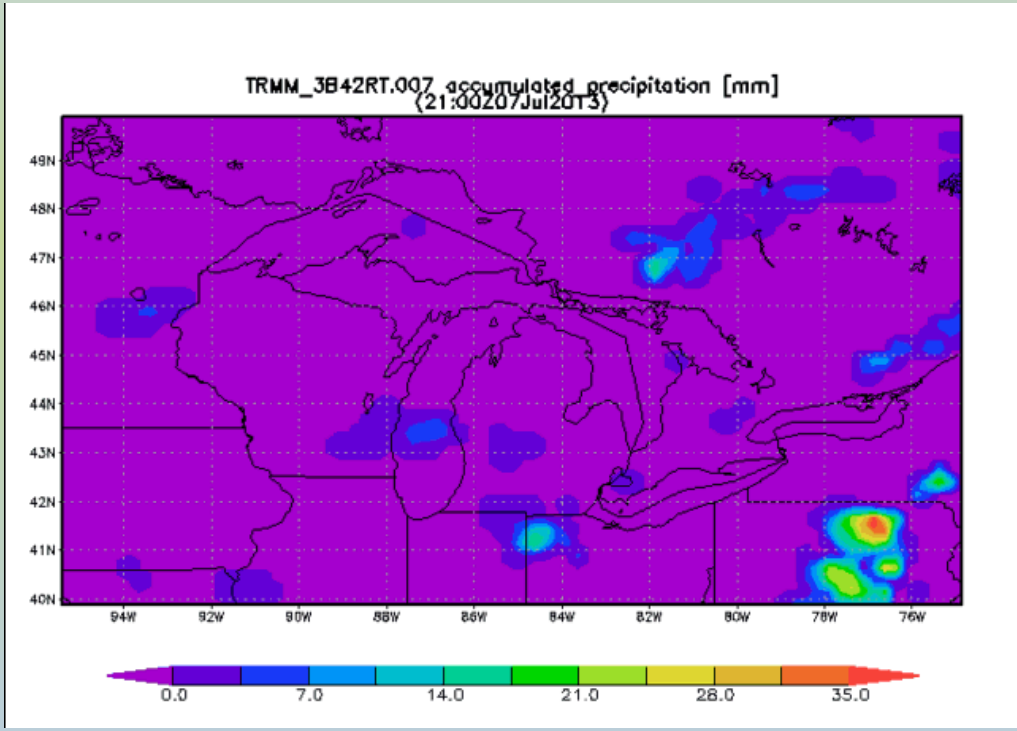
MODIS euphotic depth anomaly for Lake Michigan, June 2004



Animations

Giovanni currently offers animations as animated GIFs, which can be viewed online. The individual frames can be downloaded to create animated GIFs on a user's own system, or converted to other animation formats. Giovanni-4 will provide directly downloadable animations.

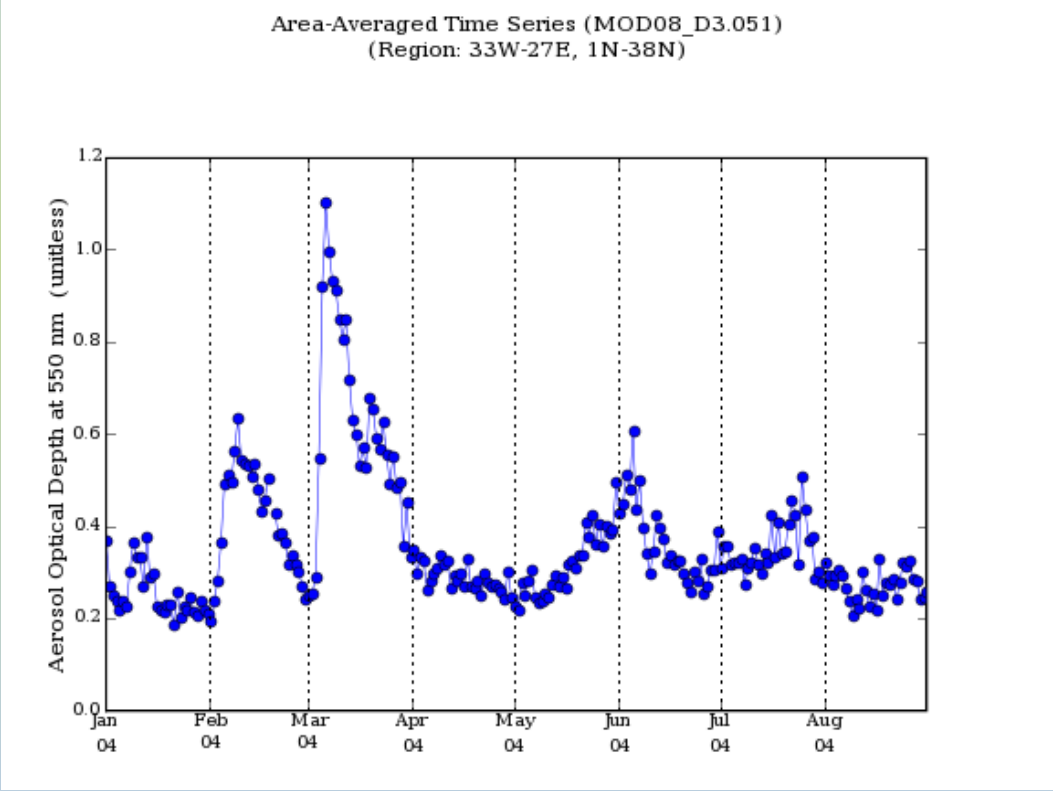
Animation of TRMM 3-Hourly Precipitation, July 9, 2013
Toronto, Canada flash flooding event



Time-series

Time-series are a powerful way to depict data trends and environmental events. Giovanni averages data from a user-selected region and plots it over a user-selected time range, returning the results in seconds (or minutes, for large areas and long time periods).

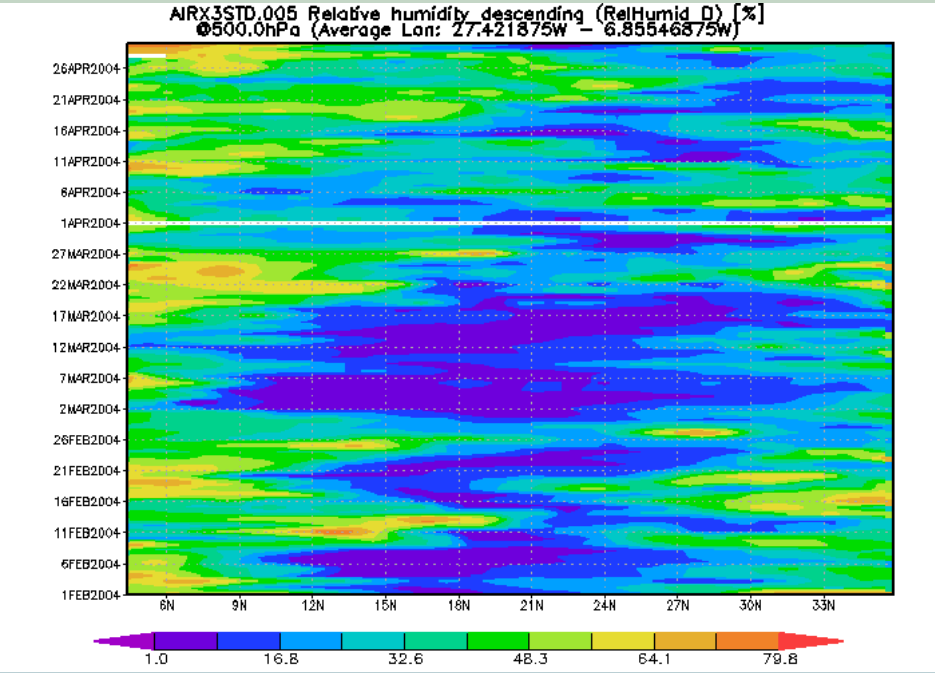
MODIS aerosol optical depth time-series off the west coast of Africa, indicating the occurrence of Saharan dust storms transported over the Atlantic Ocean.



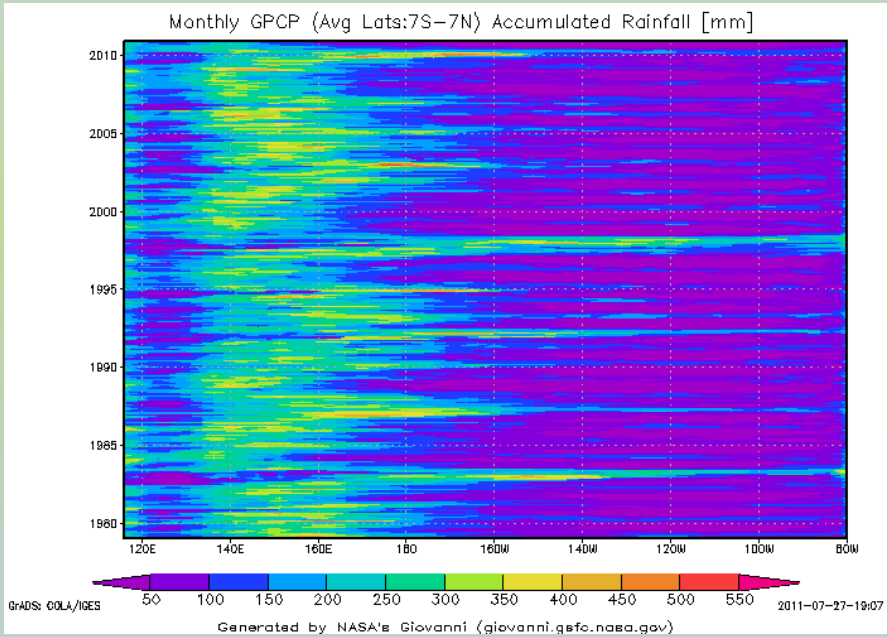
Hovmöller diagrams

Hovmoller diagrams show changes in data over latitude or longitude ranges, and are a very effective way to demonstrate the evolution of particular events through time.

Relative humidity at 500 hectoPascals over the Atlantic Ocean, February-April 2004



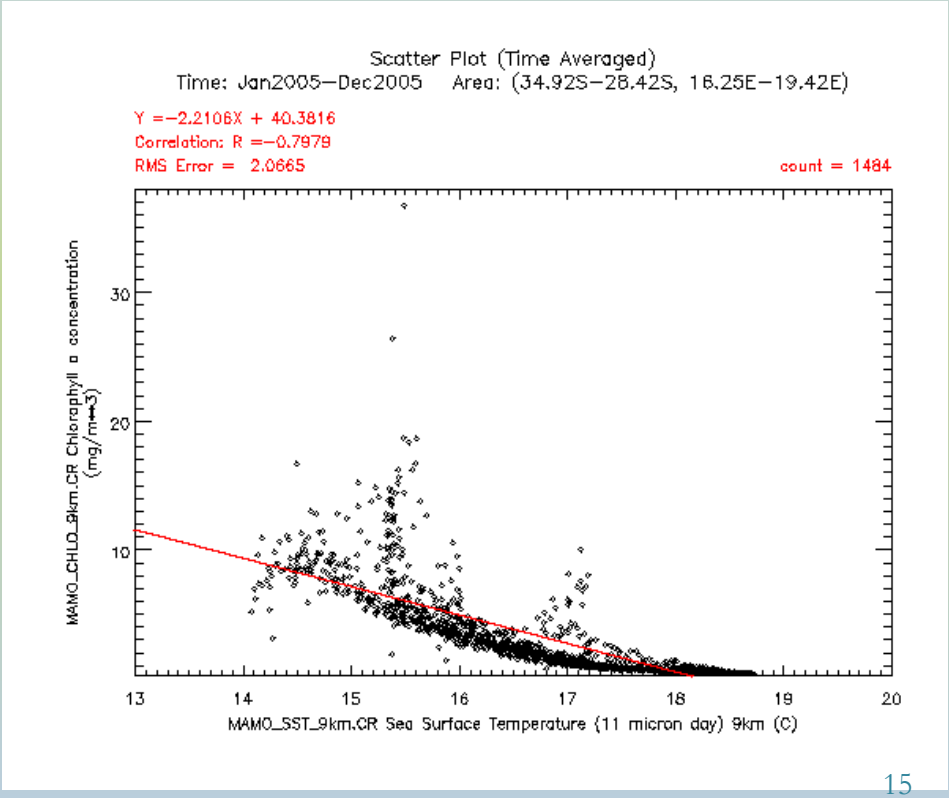
Accumulated rainfall over the Pacific Ocean, 1979-2010.



X-Y scatterplots

X-Y scatterplots directly show the relationship between two data variables in graphical form. Data variables with a strong relationship will usually have a tightly-clustered scatterplot. Data variables with a little or no relationship will have a very scattered scatterplot. Giovanni also provides the option of plotting a best-fit line to examine potential linear relationships between data variables.

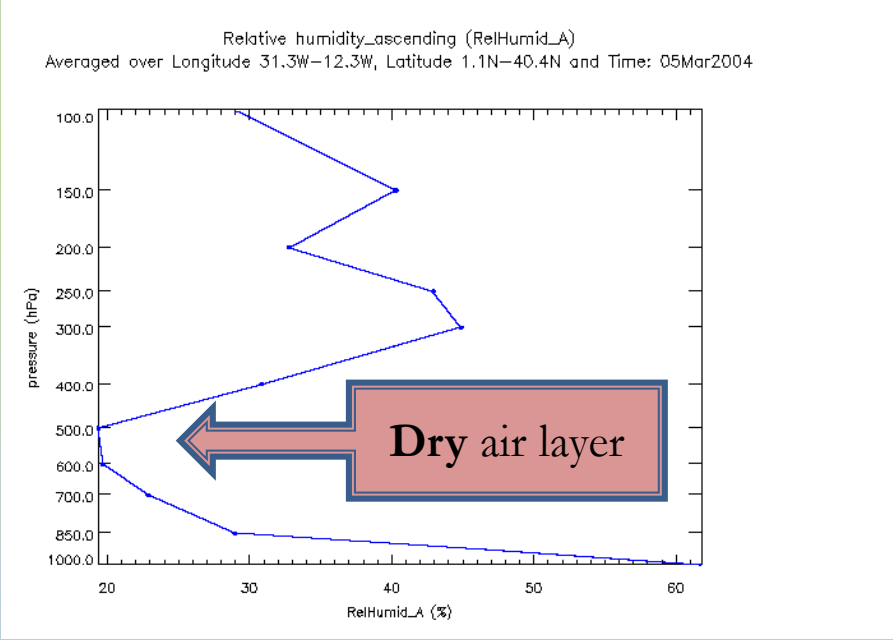
X-Y scatterplot for 2005 in the Benguela Upwelling Zone off the Southwest coast of Africa. Chlorophyll *a* concentration is plotted on the y-axis and sea surface temperature on the x-axis. The relationship between colder water and higher chlorophyll concentration is clearly portrayed in this scatterplot.



Vertical Profiles

Three-dimensional data provides cross-sections of atmospheric data from sounding instruments, such as the Atmospheric Infrared Sounder (AIRS). Vertical profile plots portray this data to give additional perspective on weather and climate processes. Model data in Giovanni provides many three-dimensional data products.

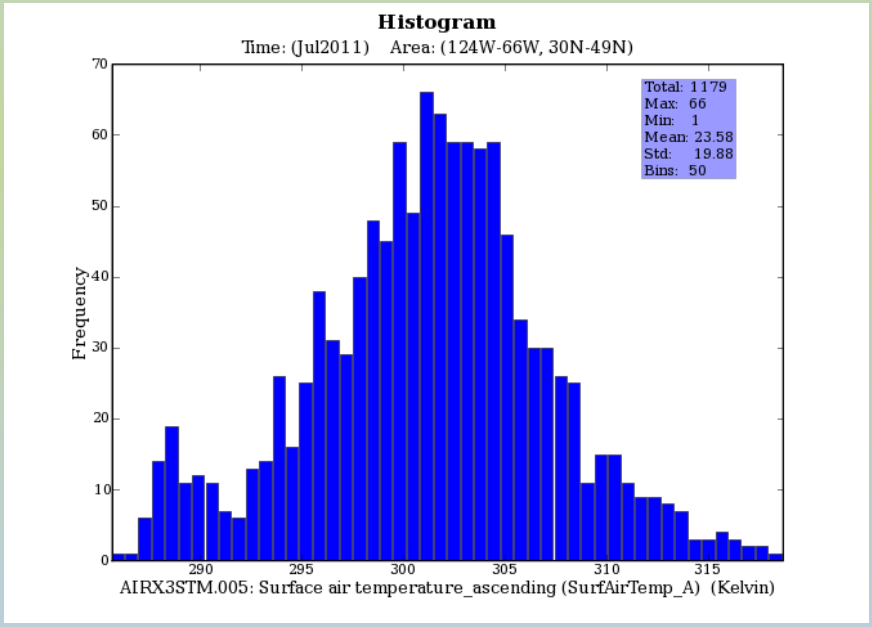
Relativity humidity data from AIRS, showing the dry Saharan air layer associated with a Saharan dust storm. Giovanni images can be easily annotated with instructive graphics.

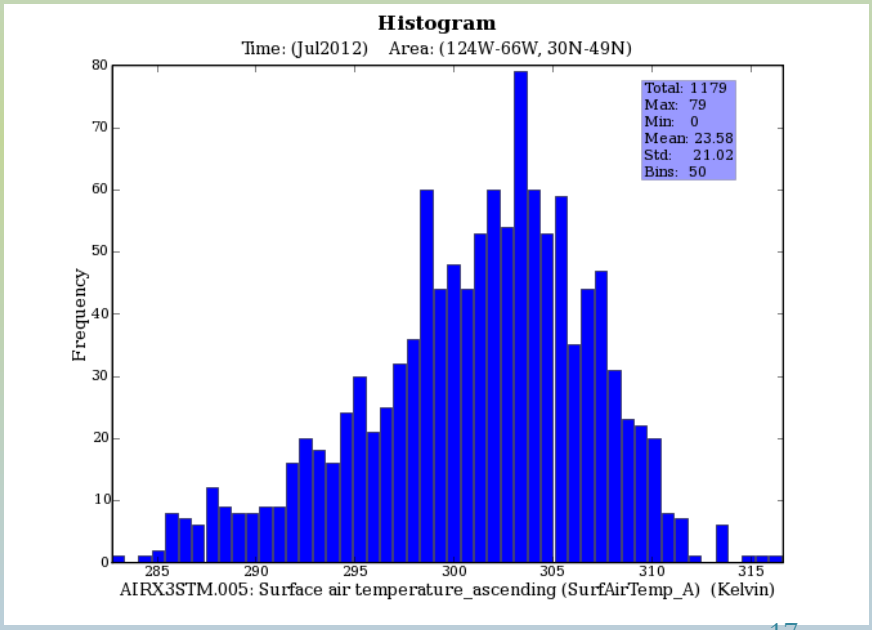


Histograms

Histograms show the distribution of data values in a selected region, and can be used effectively to data from different time-periods for the same region.

Histogram comparison of AIRS surface temperature data for July 2011 (left) and July 2012 (right) for the continental United States. A Midwest heat wave in July 2012 shifted the plot toward higher temperatures.





Part 2:
Data types in Giovanni
Useful for Public Health

There are many different data types currently in Giovanni that could be of interest to public health research. Several of these are listed below and will be described briefly in subsequent slides.

"Tier 1"

Precipitation
Temperature
Aerosol Optical Depth
Nitrogen Dioxide (NO_2)
Carbon Monoxide (CO)
Relative Humidity
Cloud Cover

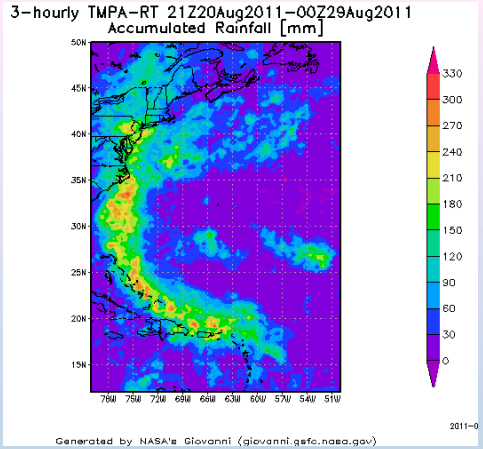
"Tier 2"

Chlorophyll concentration
Euphotic Depth
Sea Surface Temperature
Ozone (O_3)
Erythemal UV Daily Dose
NDVI/EVI
Soil Moisture

"Tier 3"

Snow Depth
Snow Mass
Snowfall Rate
Snowmelt
Fractional Snow Cover
Snow/Ice Frequency
Wind Speed
Runoff

Tier 1
Directly Useful Data Types



Precipitation:

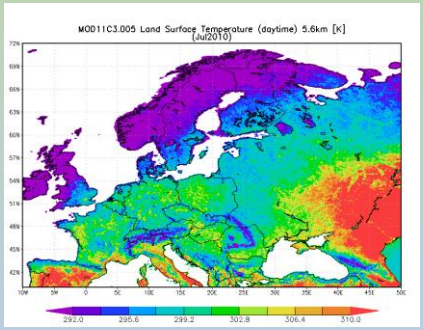
Highly correlated with waterborne diseases, insect population outbreaks, & transmission modes (i.e. shared water resources).

Giovanni has Tropical Rainfall Measuring Mission (TRMM) data products, climatological precipitation data products, and model precipitation data products

Temperature:

Fundamental variable related to water resources, drought conditions, vegetation survival, insect overwintering survival, heat stress, species ranges.

Giovanni has remotely-sensed temperature data from MODIS and AIRS, model temperature data, high-resolution temperature data for specific regions



Rebaudet S, Gazin P, Barrais R, Moore S, Rossignol E, Barthelemy N, Gaudart J, Boncy J, Magloire R, Piarroux R.
The Dry Season in Haiti: a Window of Opportunity to Eliminate Cholera.
PLOS Currents Outbreaks. 2013 Jun 10
[last modified: 2013 Jul 24]. Edition 1. doi:
10.1371/currents.outbreaks.2193a0ec4401d9526203af12e5024ddc.

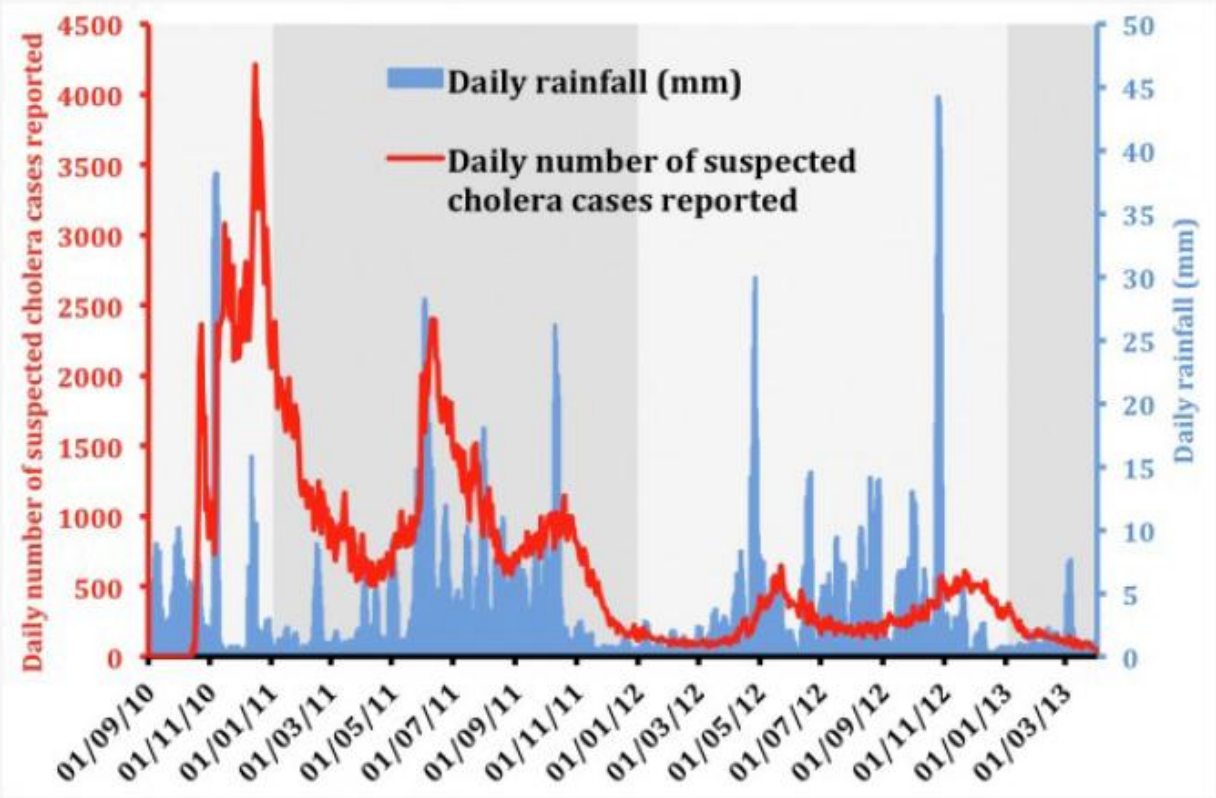


Fig. 1: Evolution of the daily suspected cholera cases and rainfall between September 2010 and March 2013.

Precipitation data were acquired from Giovanni

Radina Soebiyanto will now discuss research projects using these and other data product types available in Giovanni

INFECTIOUS DISEASE APPLICATION EXAMPLES

Malaria

- Cause:
  - *Plasmodium* spp (protozoan)
  - Carried by *Anopheles* mosquito
- Burden:
  - 250 million cases each year
  - 1 million deaths annually
  - Every 30 seconds a child dies from malaria in Africa
  - Cost ~ 1.3% of annual economic growth in high prevalence countries
- High Risk Group: Pregnant women, children and HIV/AIDS co-infection

Plasmodium infecting red blood cell

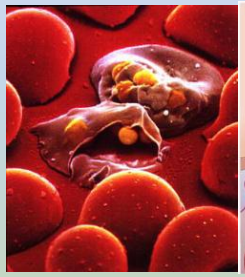


Image: Nat'l Geographic

Image: Nature

Transmission through female *Anopheles* bite

- Treatment and Prevention:

Bed nets

loor spraying

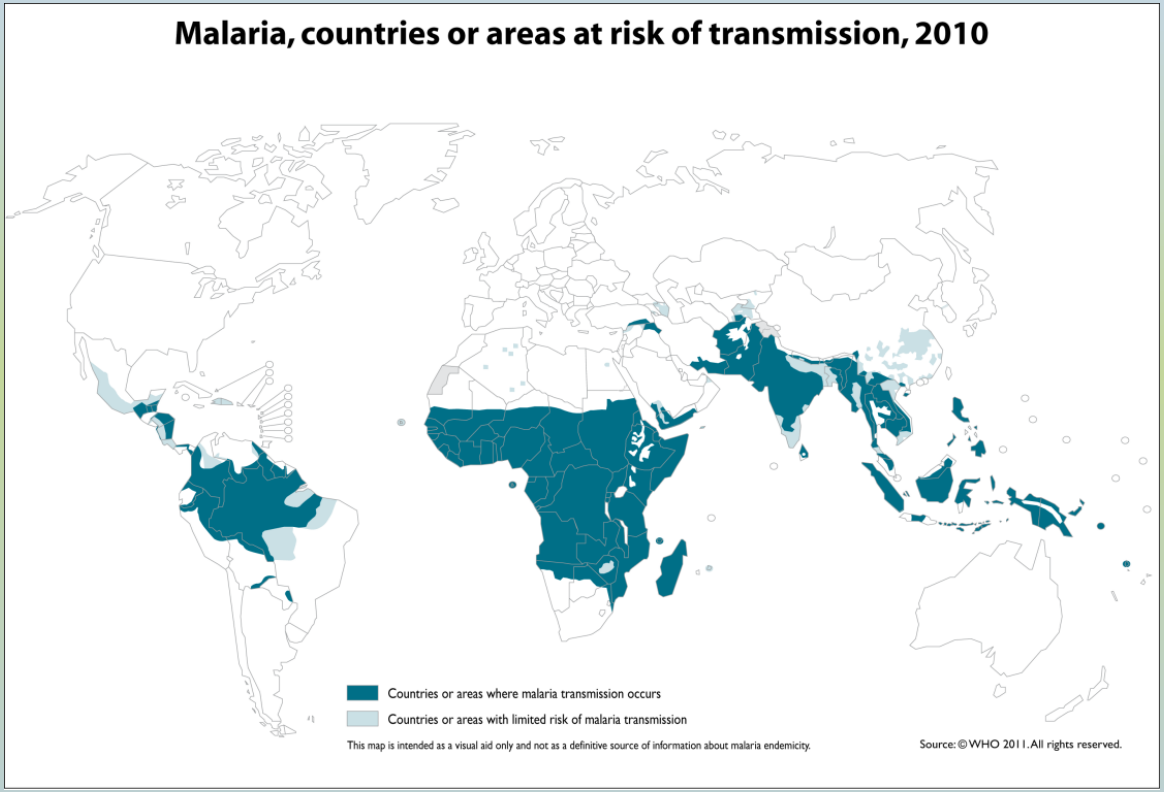
Images: WHO

Vector Control

Artemisin-based Combination Therapy

Malaria

Malaria Distribution

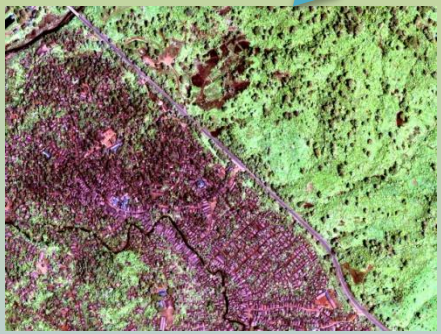


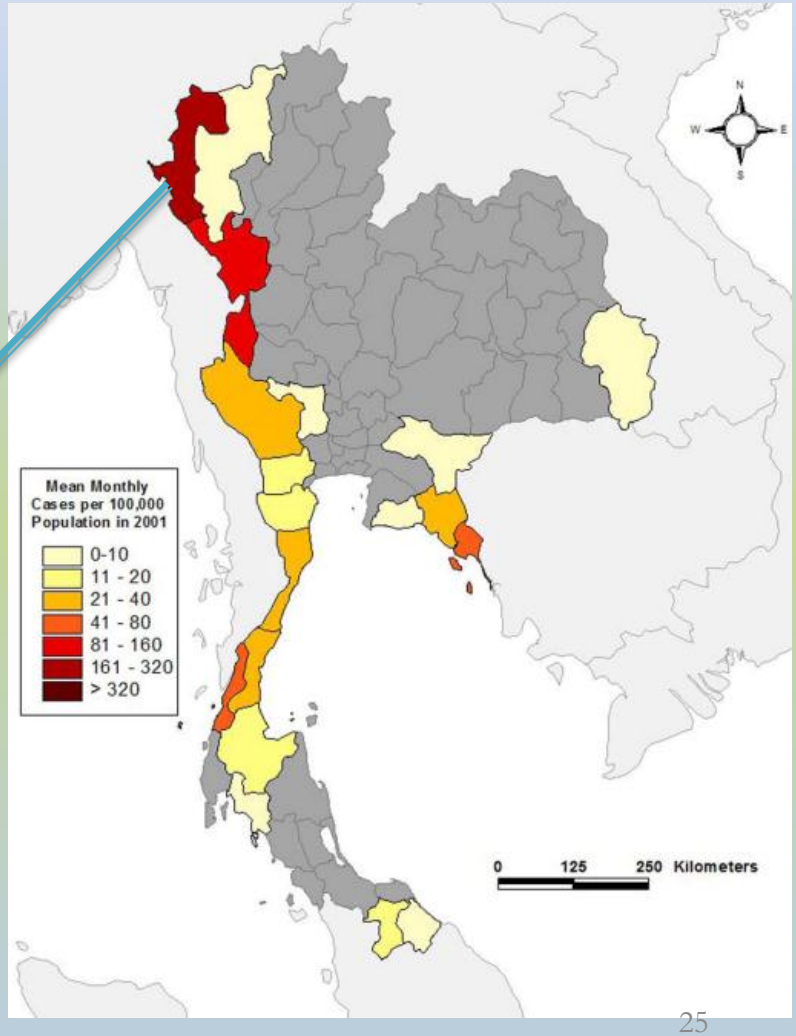
Role of climatic and environmental determinants

Determinants	Effect
Temperature	Parasite + Vector: development and survival
Rainfall	Vector breeding habitat
Land-use, NDVI	Vector breeding habitat
Altitude	Vector survival
ENSO	Vector development, survival and breeding habitat

Malaria in Thailand

- Leading cause of morbidity and mortality in Thailand
- ~50% of population live in malarious area
- Most endemic provinces are bordering Myanmar & Cambodia
  - Significant immigrant population
  - Mae La Camp
    - Largest refugee camp
    - >30,000 population

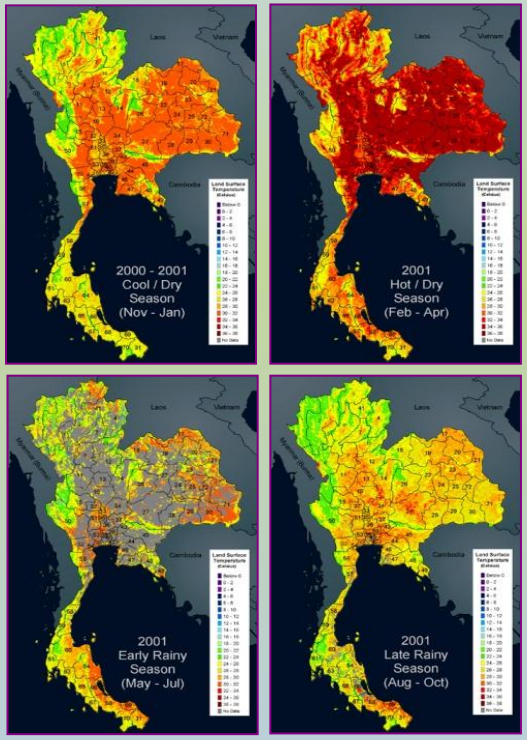




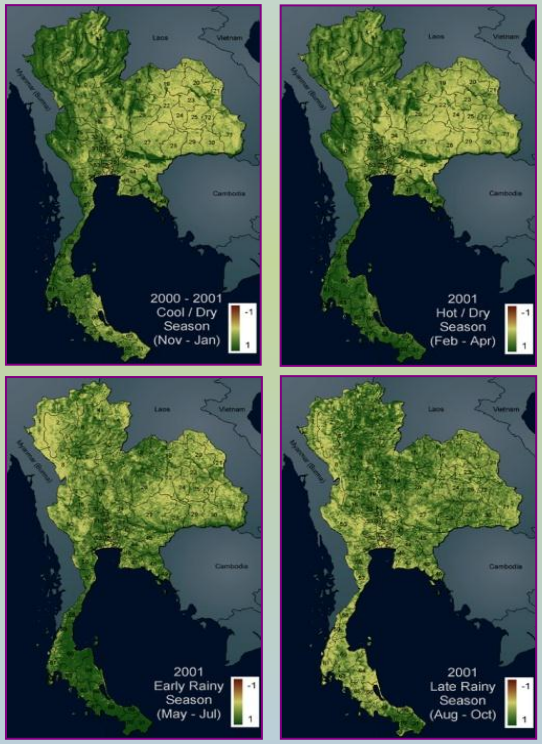
Malaria in Thailand

- Satellite-observed meteorological & Environmental Parameters for 4 Thailand seasons

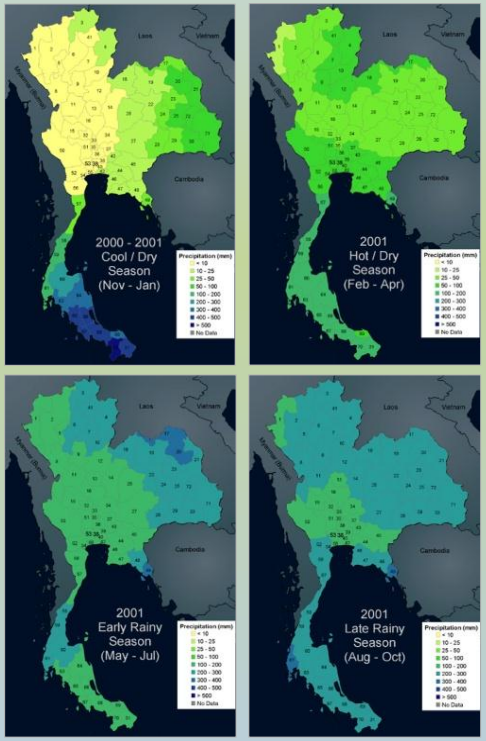
Surface Temperature
MODIS Measurements



Vegetation Index
AVHRR & MODIS Measurements



Rainfall
TRMM Measurements

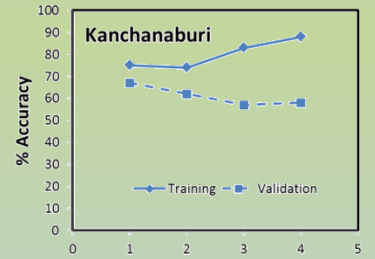


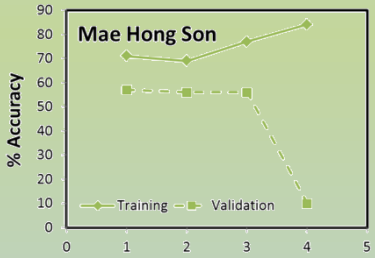
Malaria in Thailand

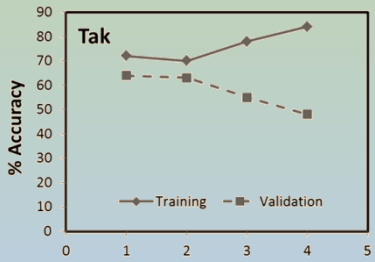
- Neural Network training and validation accuracy

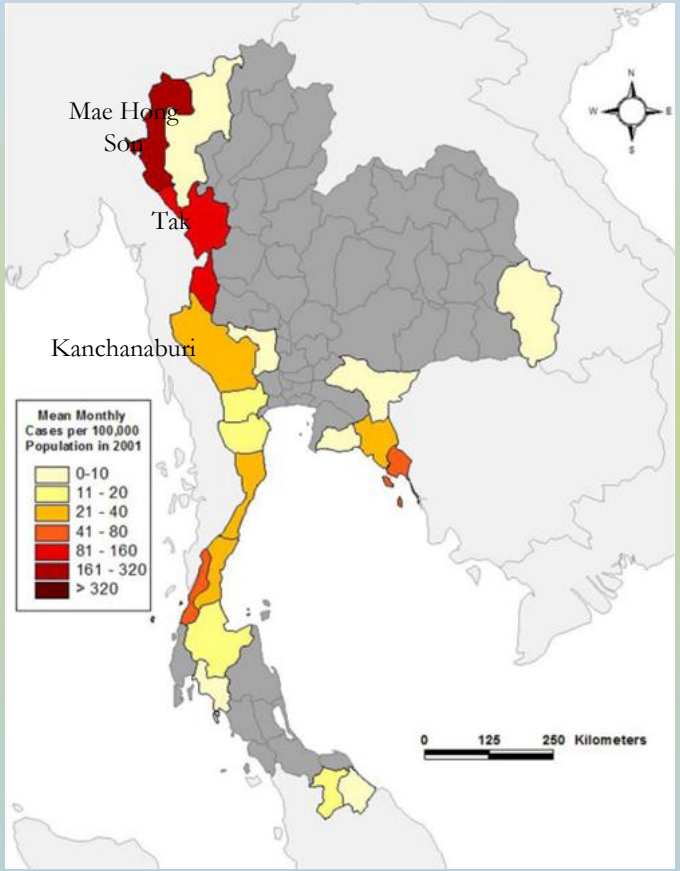
	Input	Hidden Layer	Hidden Node
Model 1	t, T, P, P (lag 1), H, V	1	1
Model 2	t, P, P (lag 1), H, V	1	1
Model 3	t, T, P, P (lag 1), H, V	1	2
Model 4	t, T, P, P (lag 1), H, V	1	3

t = time, T = temperature, P = precipitation, H = humidity, V = NDVI



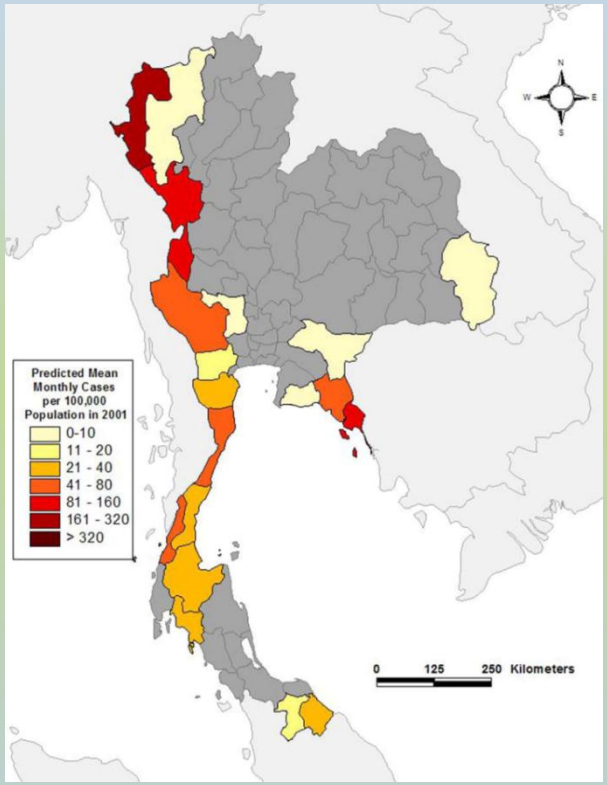




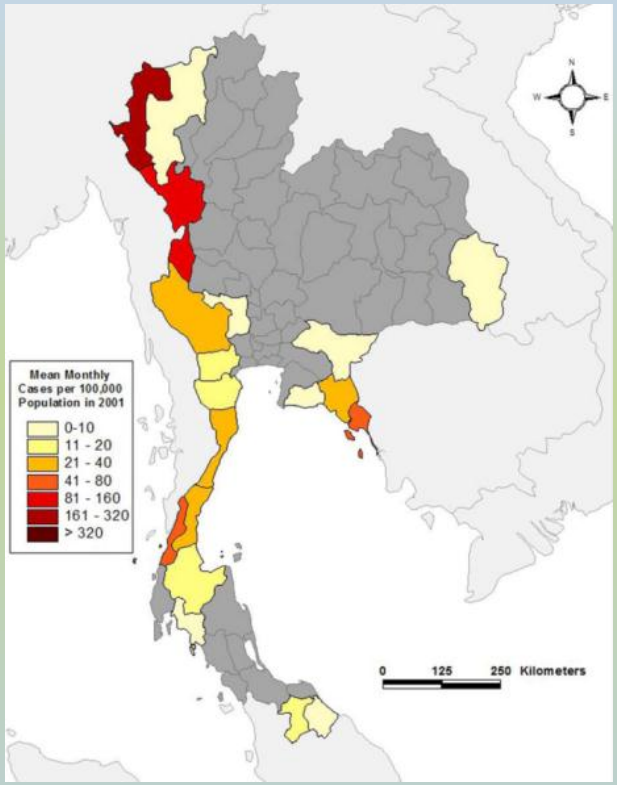


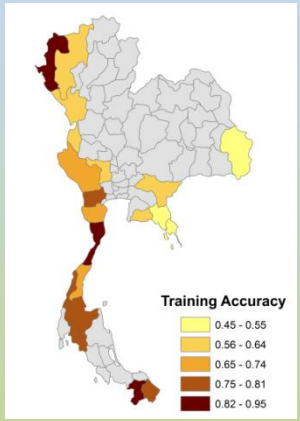
Malaria in Thailand

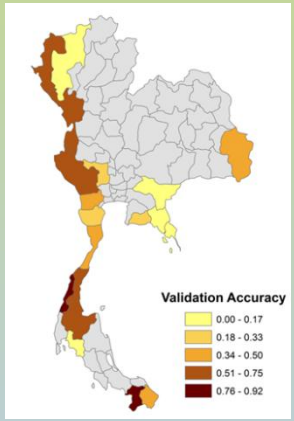
Hindcast Incidence



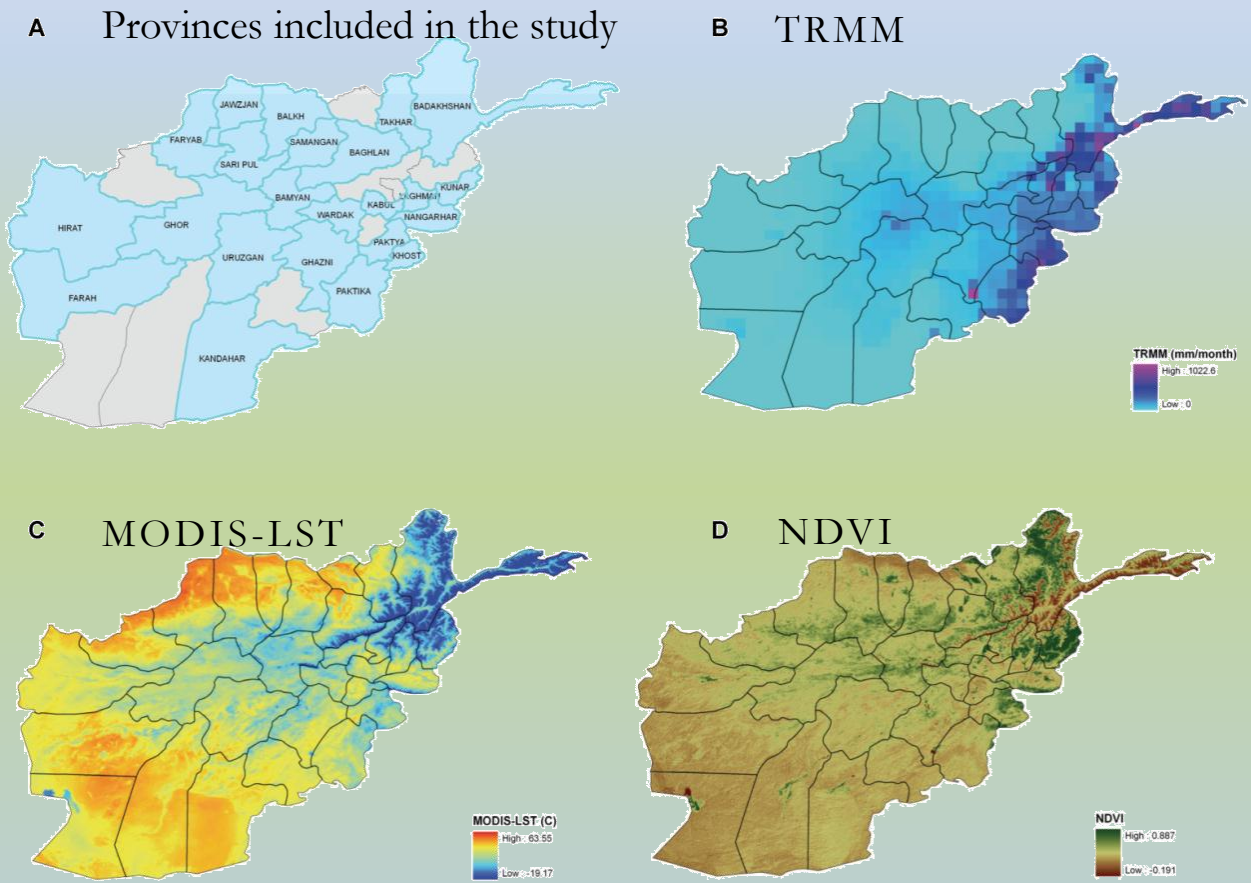
Actual Malaria Incidence

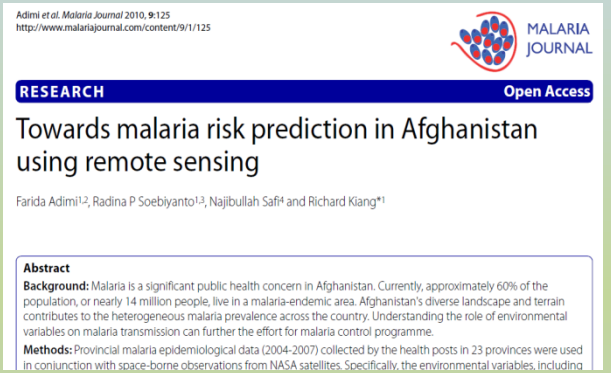






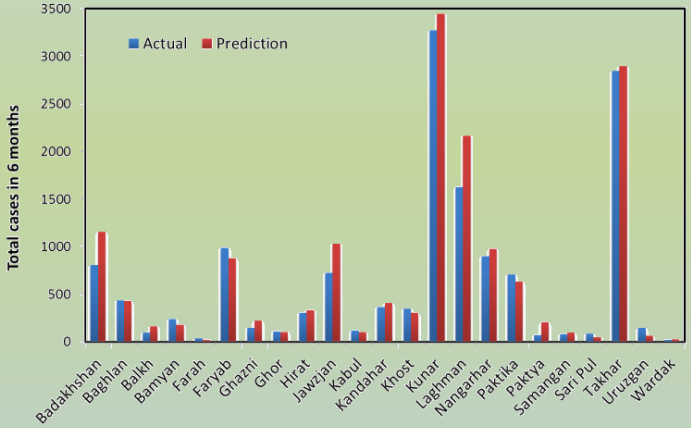
Malaria in Afghanistan

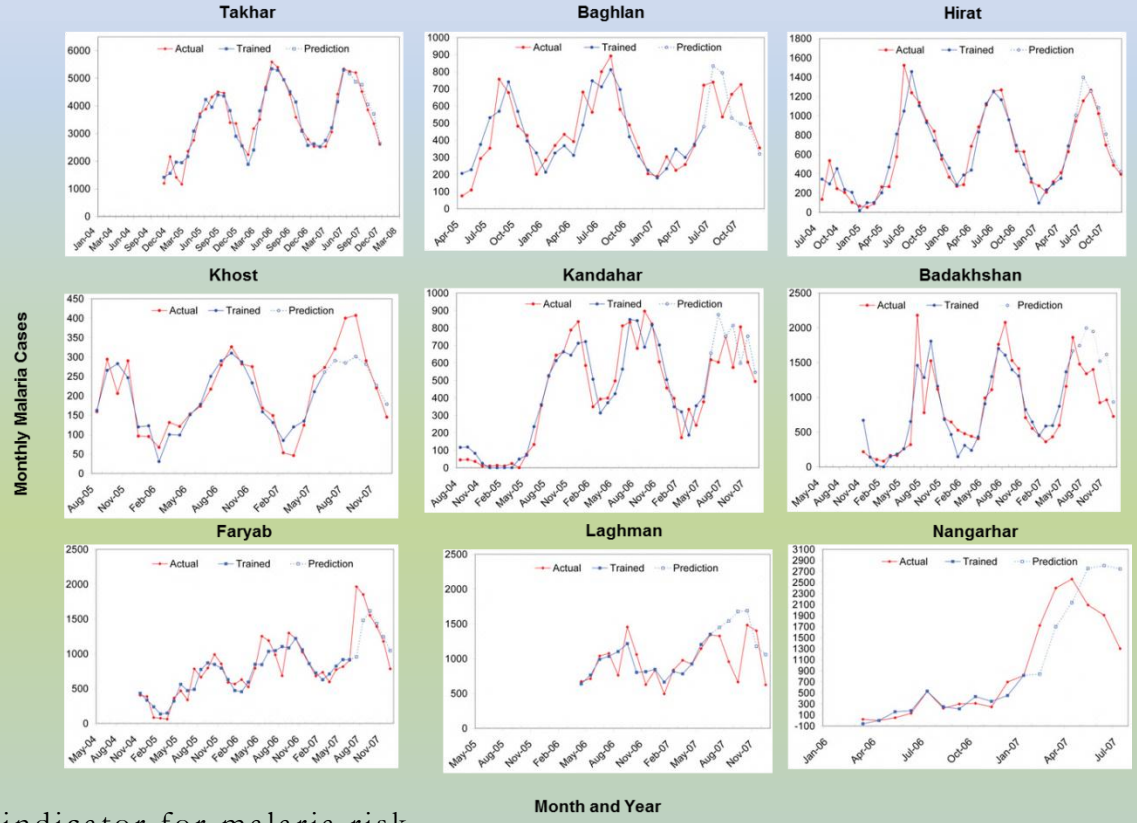




Adimi *et al. Malaria Journal* 2010, **9**: 125

Malaria in Afghanistan

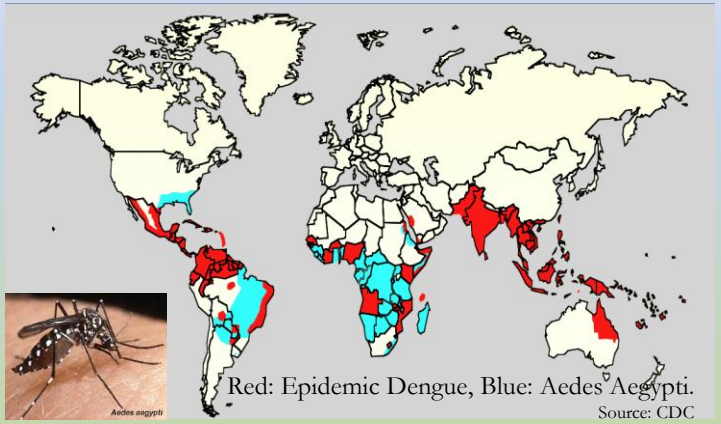


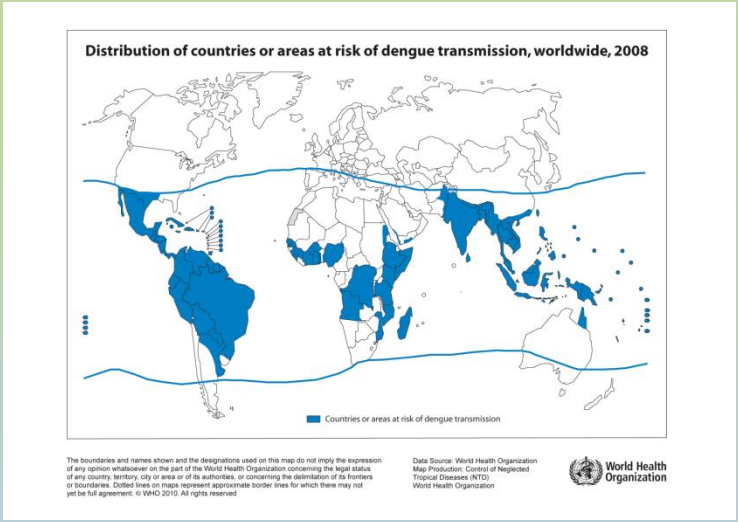


- NDVI and temperature were a strong indicator for malaria risk
- Precipitation is not a significant factor → Malaria risk is mainly due to irrigation as implied from the significant contribution from NDVI
- Average R^2 is 0.845
- Short malaria time series (<2 years) pose a challenge for modeling and prediction

Dengue

- Endemic in more than 110 countries
  - Tropical, subtropical, urban, peri-urban areas
- Annually infects 50 – 100 million people worldwide
- 12,500 – 25,000 deaths annually
- Symptoms: fever, headache, muscle and joint pains, and characteristic skin rash (similar to measles)
- Primarily transmitted by *Aedes* mosquitoes
  - Live between 35°N - 35°S latitude, >1000m elevation
- Four serotypes exist
  - Infection from one serotype may give lifelong immunity to that serotype, but only short-term to others
  - Secondary infection increases the severity risk

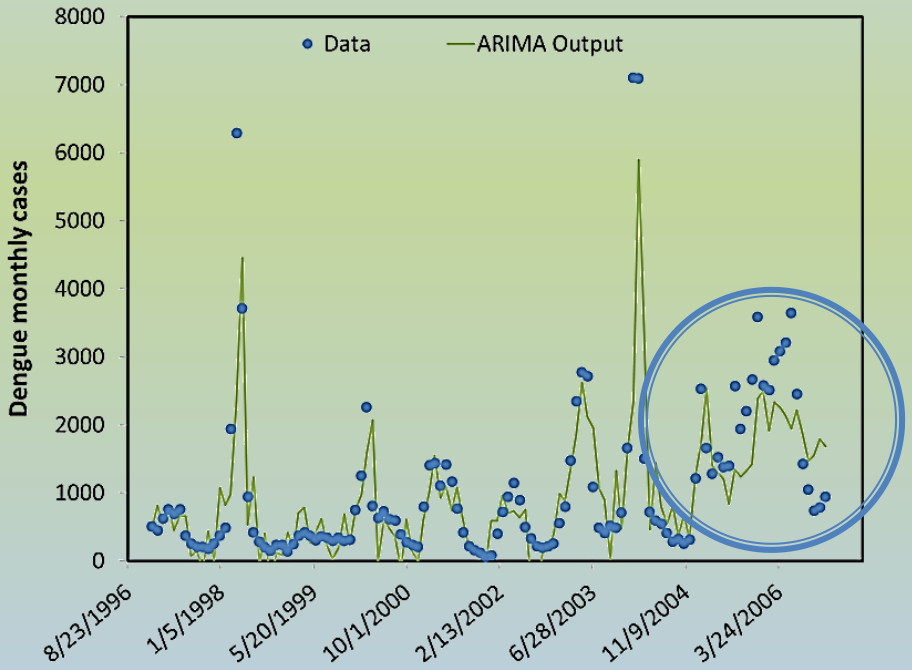




Dengue in Indonesia

- Environmental variables used
  - Temperature, dew point, wind speed, TRMM, NDVI
- Modeling method
  - ARIMA – Auto Regressive Integrated Moving Average
  - Classical time series regression
  - Accounts for seasonality

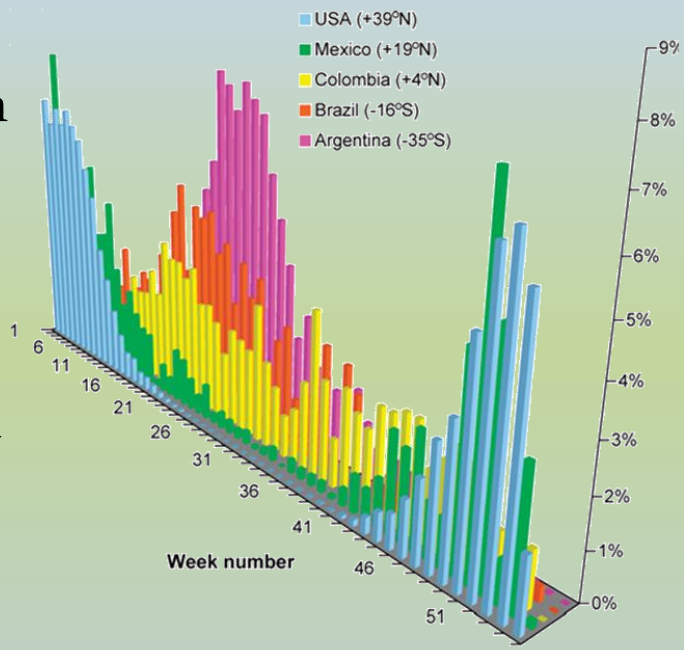
- Result
  - Best-fit model uses TRMM and Dew Point as inputs
  - Peak timing can be modeled accurately up to year 2004
  - Vector control effort by the local government started in the early 2005



Seasonal influenza

- Worldwide annual epidemic
  - Infects 5 – 20% of population with 500,000 deaths
- Economic burden in the US ~US$87.1billion
- Spatio-temporal pattern of epidemics vary with latitude
  - Role of environmental and climatic factors
- Temperate regions: distinct annual oscillation with winter peak
- Tropics: less distinct seasonality and often peak more than once a year

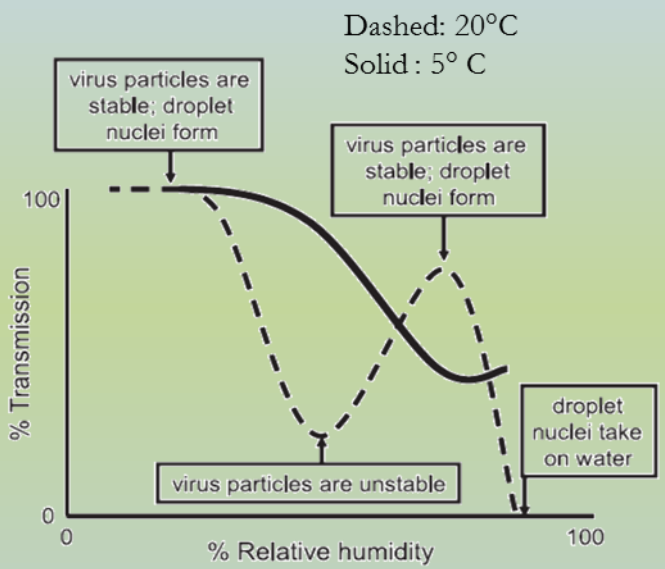
Viboud et al. (2006), PLoS Med. 3(4):e89



Seasonal influenza

- Factors implicated in influenza

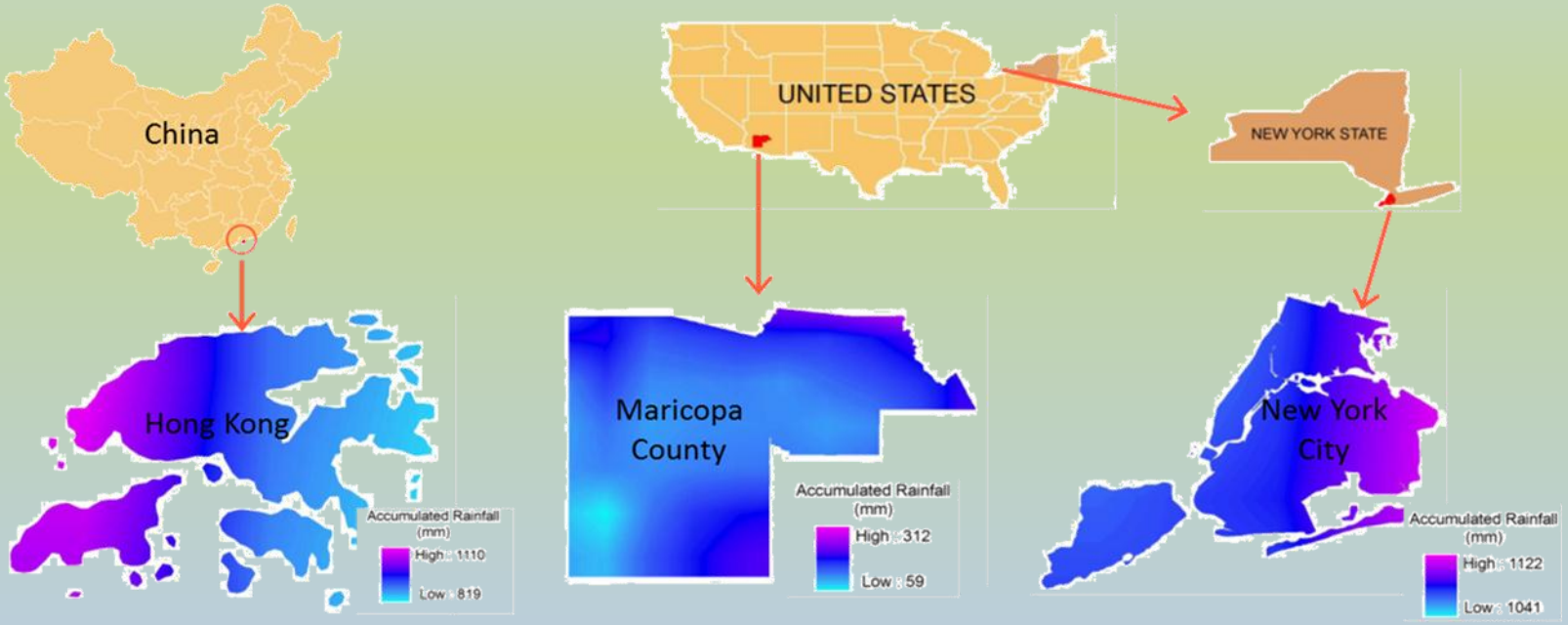
Influenza Process	Factors	Relationship
Virus Survivorship	Temperature	Inverse
	Humidity	Inverse
	Solar irradiance	Inverse
Transmission Efficiency	Temperature	Inverse
	Humidity	Inverse
	Vapor pressure	Inverse
	Rainfall	Proportional
	ENSO	Proportional
	Air travels and holidays	Proportional
Host susceptibility	Sunlight	Inverse
	Nutrition	Varies

- *Ex Vivo* study showing efficient transmission at dry and cold condition *[Lowens et al., 2007]*
- 

- High temperature (30°C) blocks aerosol transmission *but not contact transmission*

Seasonal influenza

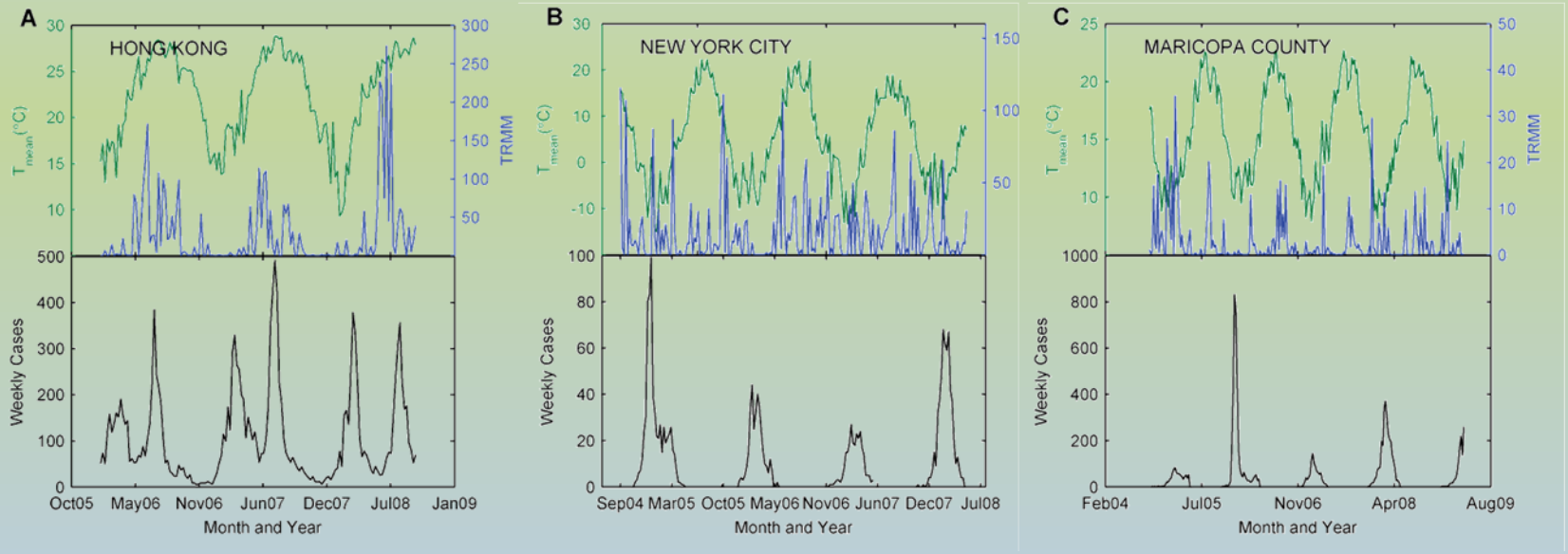
	Hong Kong, China	Maricopa County, AZ	New York City, NY
Center Lat.	22° N	33° N	40° N
Climate	Sub-Tropical	Sub-Tropical	Temperate
General Condition	Hot & humid during summer. Mild winter, average low of 6°C	Dry condition. Mean winter low is 5°C, and summer high is 41°C	Cold winter, average low of -2°C. Mean summer high is 29°C



Seasonal influenza

DATA

- Weekly lab-confirmed influenza positive
- Daily environmental data were aggregated into weekly
- Satellite-derived data
  - TRMM 3B42
  - LST - MODIS
- Ground station data



Seasonal influenza

- Several techniques were employed, including:

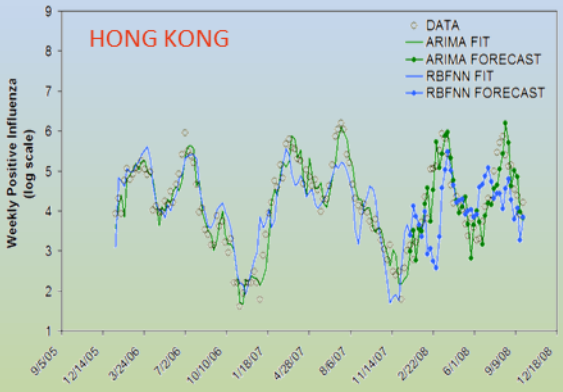
***ARIMA* (AutoRegressive Integrated Moving Average)**

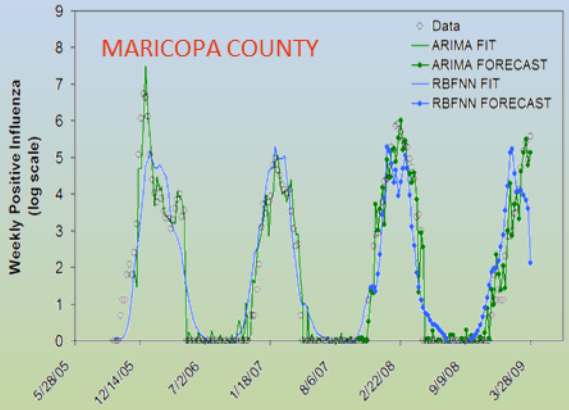
- Classical time series regression Accounts for autocorrelation and seasonality properties
- Climatic variables as covariates
- Previous week(s) count of influenza is included in the inputs
- Results published in PLoS ONE 5(3): 9450, 2010

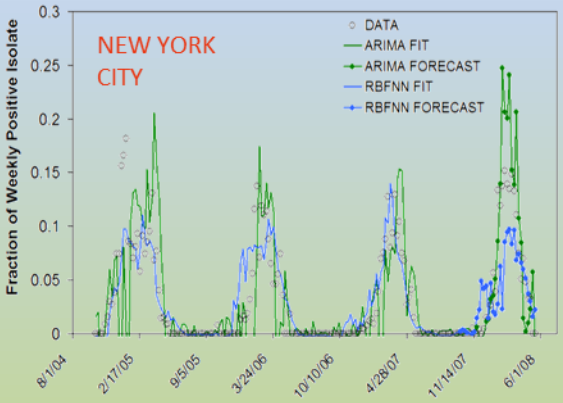
***Neural Network* (NN)**

- Artificial intelligence technique
- Widely applied for
  - approximating functions,
  - Classification, and
  - pattern recognition
- Takes into account nonlinear relationship
- Radial Basis Function NN with 3 nodes in the hidden layer
- Only climatic variables and their lags as inputs/predictors

Seasonal influenza



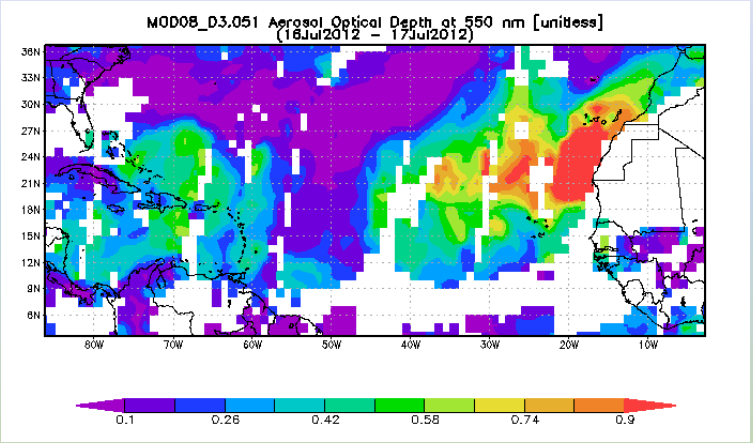




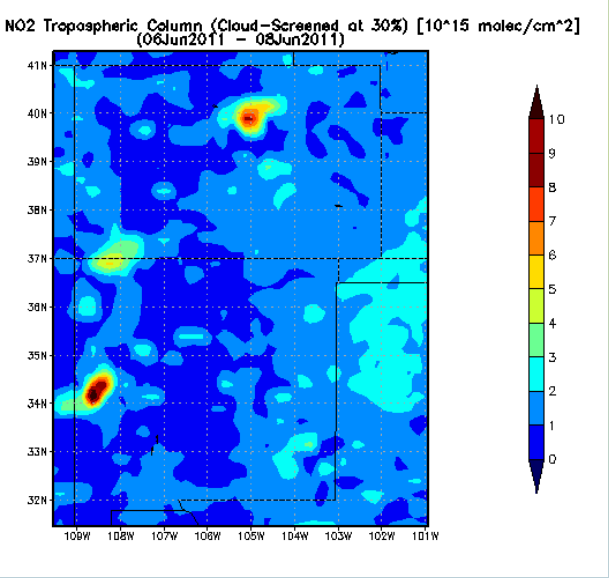
- NN models show that ~60% of influenza variability in the US regions can be accounted by meteorological factors
- ARIMA model performs better for Hong Kong and Maricopa
  - Previous cases are needed
  - Suggests the role of contact transmission
- Temperature seems to be the common determinants for influenza in all regions

Air Quality data types

Aerosol Optical Depth:
Relevant to wildfire smoke, dust storms, remobilization of contaminants in soils and vegetation, urban air pollution,



Nitrogen Dioxide (NO_2):
Direct product of combustion; indicates location of fires, urban pollution levels – can be used to examine commuting, energy production on a daily basis



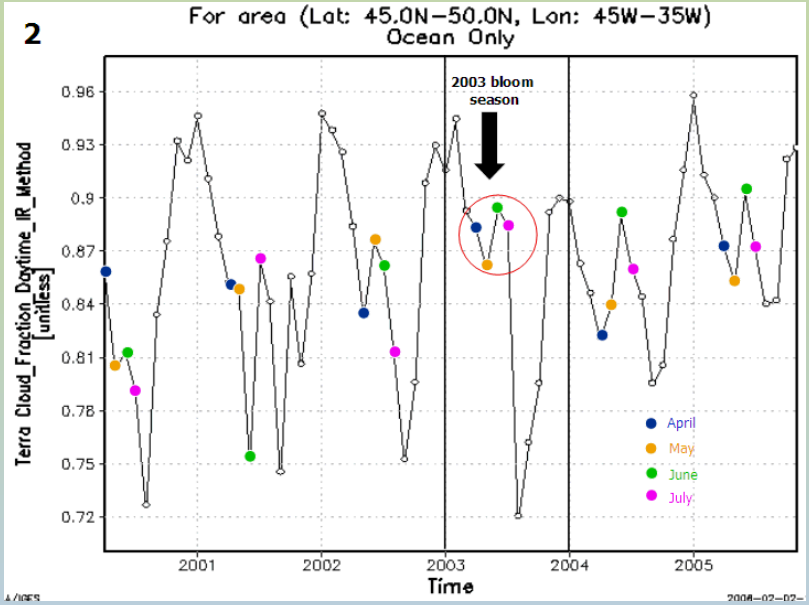
Carbon Monoxide (CO):
By-product of combustion; also can be used to examine air quality impact of fires, urban air pollution levels

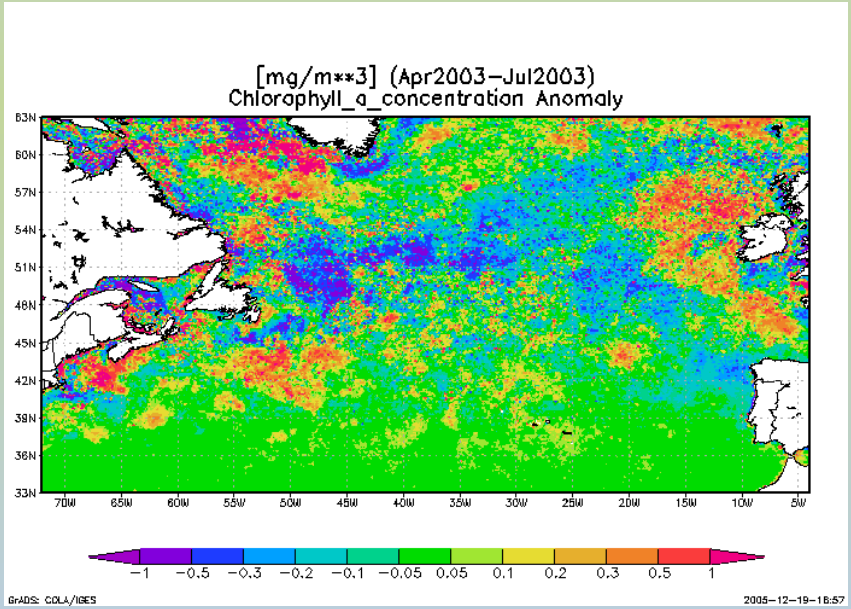
PM 2.5
EPA data product of particulate matter (USA only)

Meteorological Indicators

Relative Humidity:
Indicator of heat stress potential, meteorological environment, shifts in weather patterns, insect (vector) survival, transmission efficiency

Cloud Cover:
General indicator of overall meteorological conditions, rainfall potential, drought conditions, weather patterns, flash flooding, anomalous seasonal conditions

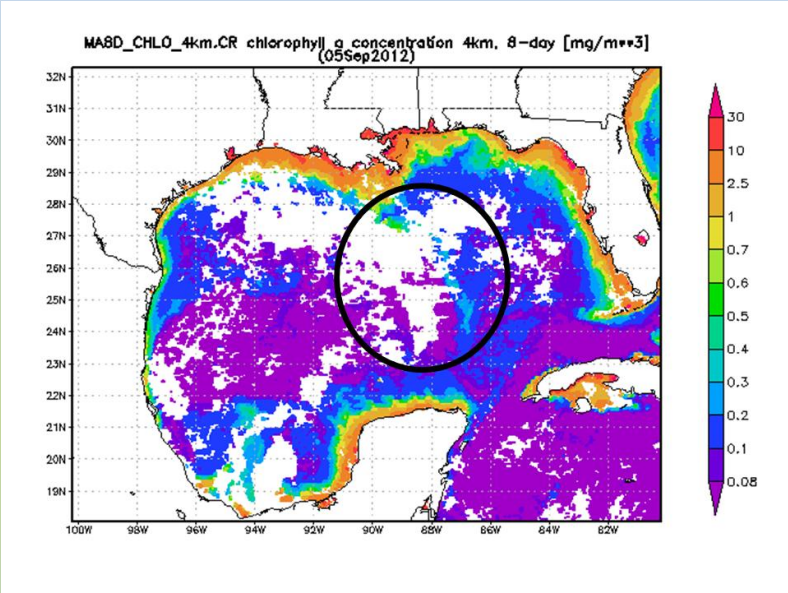




Tier 2: Indirectly Useful Data Types

WATER QUALITY

Chlorophyll concentration:
Shows the occurrence of phytoplankton populations and growth – can be related to waterborne diseases like cholera, seafood contamination ("red tides"), fish mortality, severe storm effects. Also related to fishery success or failure.



Euphotic Depth:
Direct indicator of water clarity, related to storm runoff, pollutant transport, transport of disease vectors and organisms, recreation impact (beach closure)

Sea Surface Temperature:
Important for phytoplankton growth, storm occurrence, regional rainfall, "teleconnections" with weather patterns around the world

Tier 2
Indirectly Useful Data Types

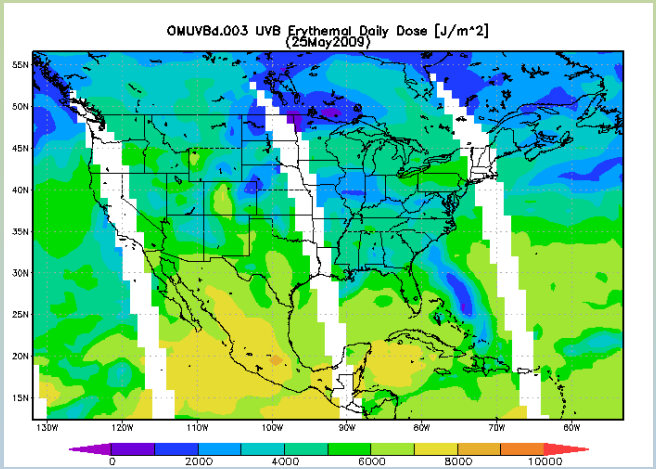
Ozone (O_3)

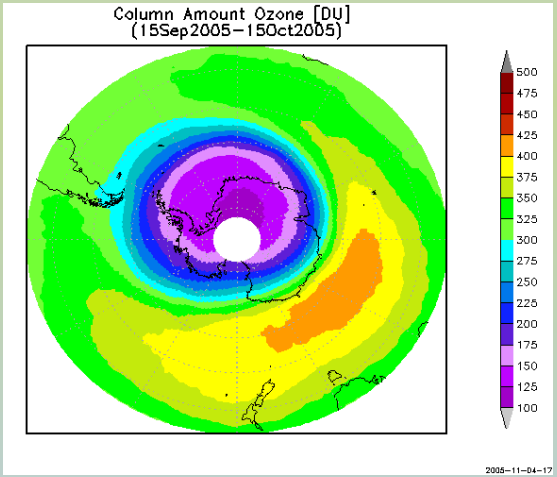
Tropospheric ozone: Air pollution indicator, related to oxidation of NOx, airborne organics; trigger for air quality alerts

Stratospheric ozone: Related to solar ultraviolet (UV) radiation transmission; potential carcinogenic and mutagenic agent

Erythemal UV Daily Dose:

Measurement of ground level UV radiation exposure



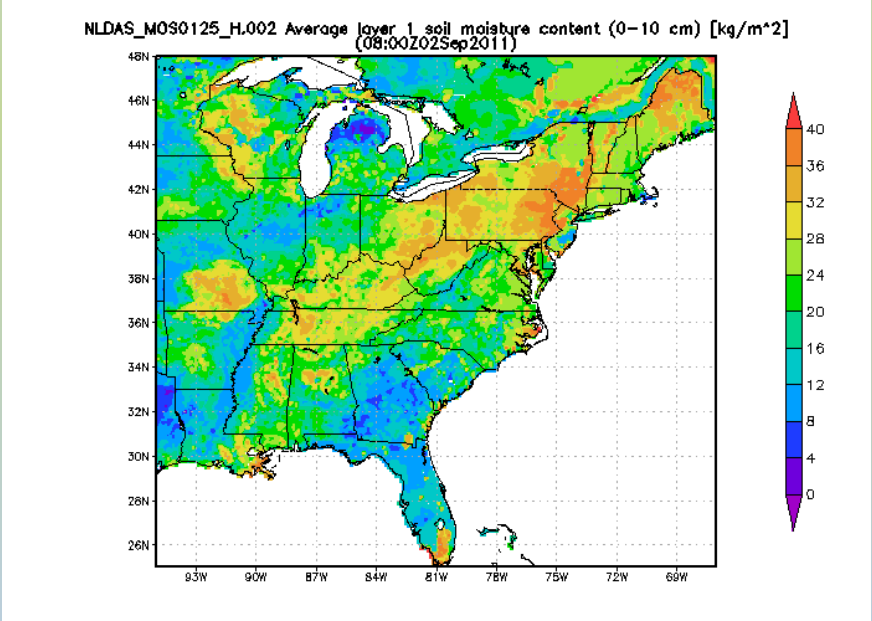


Tier 2
Indirectly Useful Data Types

Normalized Difference Vegetation Index (NDVI)
Enhanced Vegetation Index (EVI)
Related to rainfall, drought and "wet" conditions, insect (vector) life cycles, crop survival

Soil Moisture
Related to rainfall, drought conditions, wetland status, insect (vector) life cycle, irrigation needs

Hourly NLDAS surface soil moisture during passage of Hurricane Lee

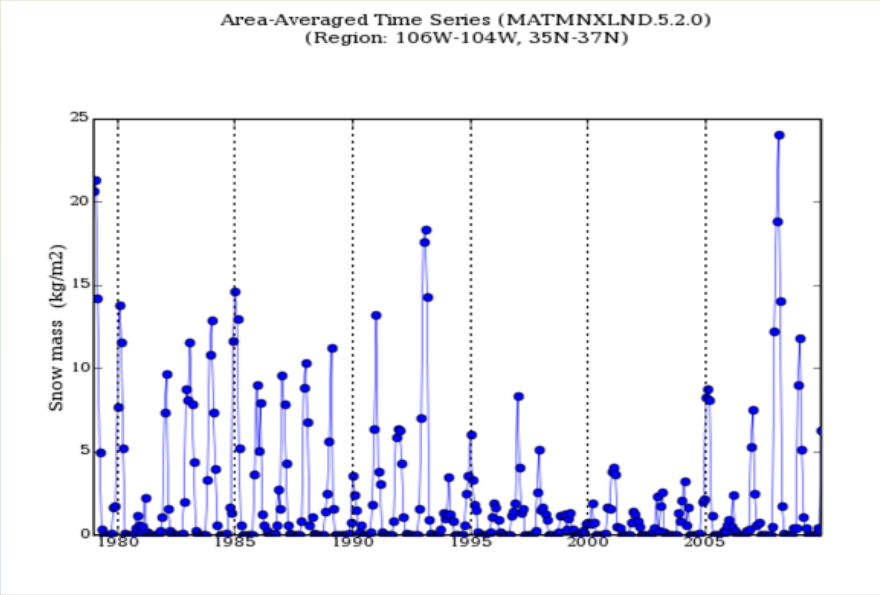


Tier 3
Potentially Useful Data Types

Snow Depth; Snow Mass; Snowfall Rate; Snowmelt; Fractional Snow Cover; Snow/Ice Frequency

All of these data types are related to water resource availability, particularly crucial during drought conditions. Snow and ice can be leading indicators of short-term and long-term climate shifts. Snowmelt can also be indicative of major spring flood potential.

Northern New Mexico
Snow Mass time-series
1979-2010



Tier 3
Potentially Useful Data Types

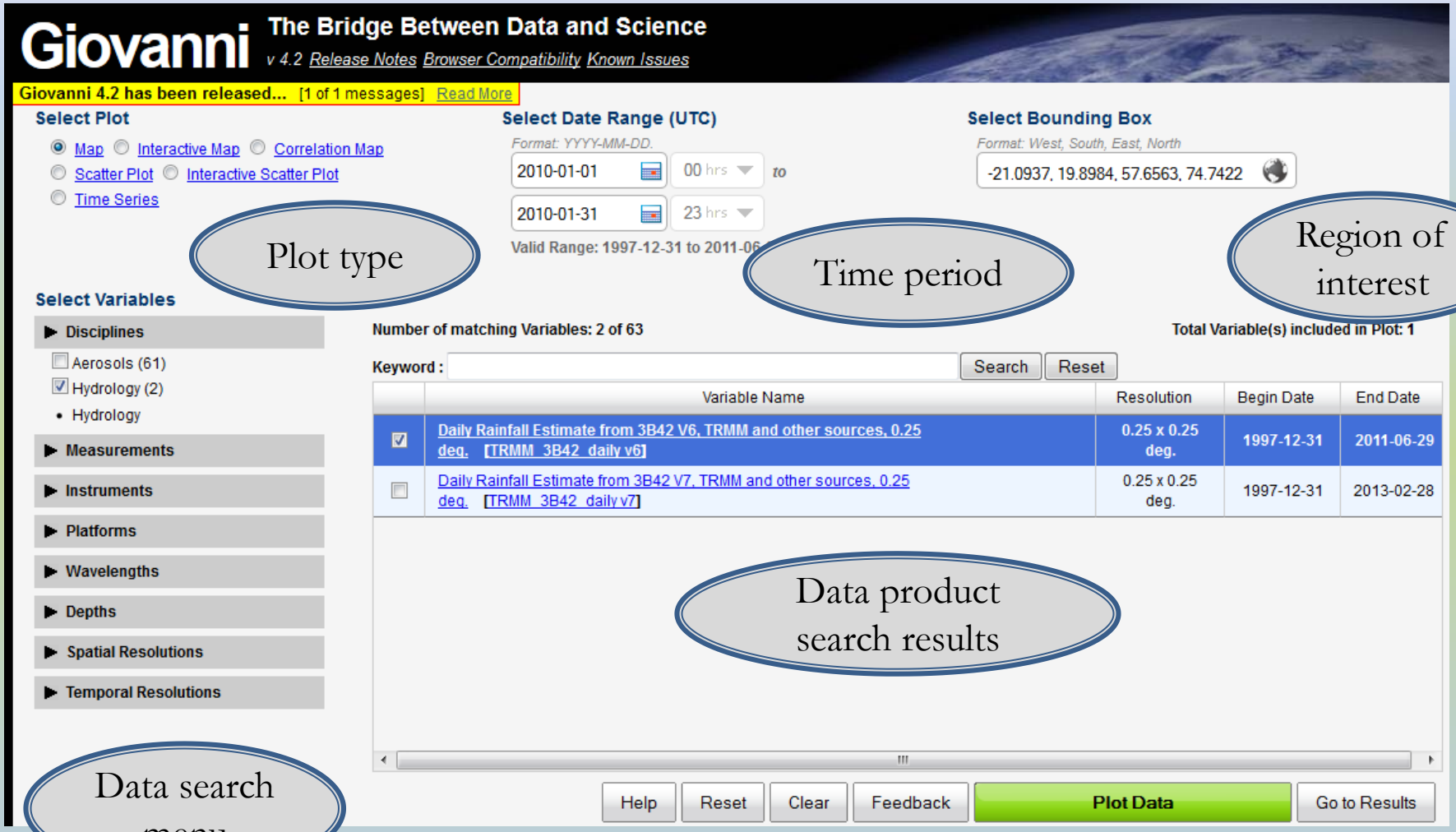
Wind Speed and Wind Direction:
Indicator of the potential for transport of air pollution and disease vectors

Runoff:
Indicator of rainfall intensity, snowmelt effects, flood potential, transport of water pollution, transport of waterborne nutrients contributing to eutrophication in lakes, bays, and coastal waters, transport of waterborne diseases

Questions and “hands-on” time

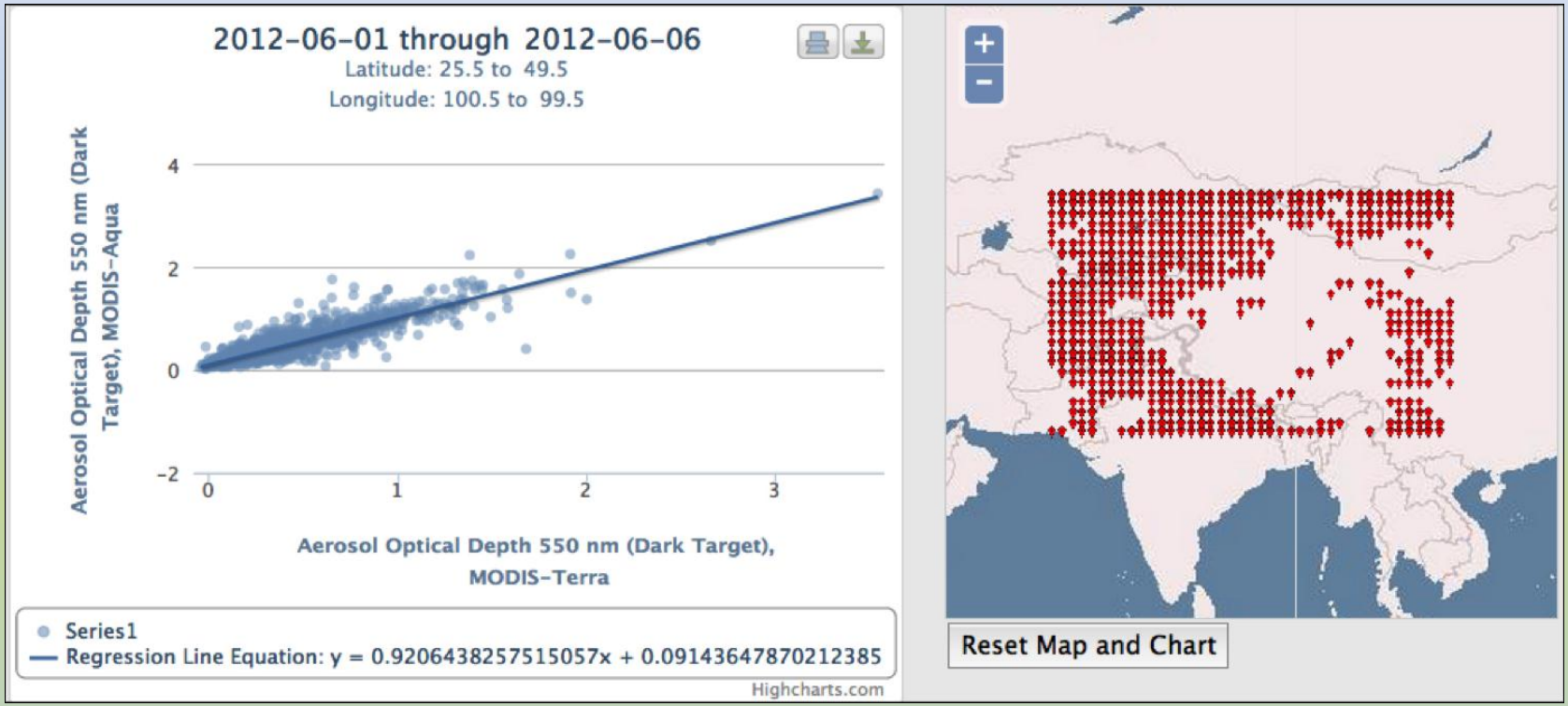
Giovanni-4 :
a very brief introduction

The new Giovanni-4 data interface



Giovanni-4 will add:

Interactive X-Y scatterplot



Giovanni-4 also provides an improved method of saving data selection and plot criteria, so that Giovanni analysis sessions can be saved and shared.

Summary

From the beginning, Giovanni development and implementation has emphasized rapid analytical results and a variety of easily-manipulated data visualizations. This focus has made it a very popular scientific research tool.

Giovanni-4 will maintain these capabilities, enhanced with a simpler user interface, more visualization options, and faster generation of results.