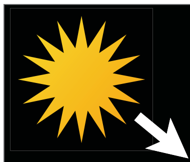
Satellite-derived datasets

Short Course on Climate Datasets and Tools
AMS Annual Meeting 2010

Eric Bruning
University of Maryland / ESSIC
Cooperative Institute for Climate and Satellites

Outline

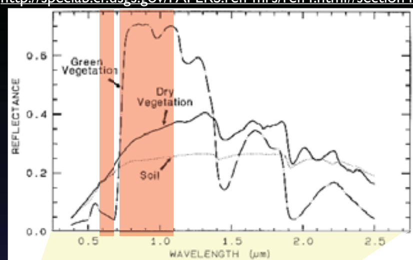
- Satellite remote sensing and retrieval methods. Calibration.
- Satellite datasets by topical area
- Strategies for data wrangling
- Example: combining disparate datasets



What does a satellite measure?

Example: Detection of vegetation (NDVI)

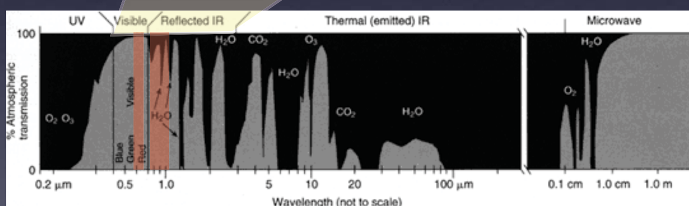
http://speclab.cr.usgs.gov/PAPERS.refl-mrs/refl4.html#section4.3



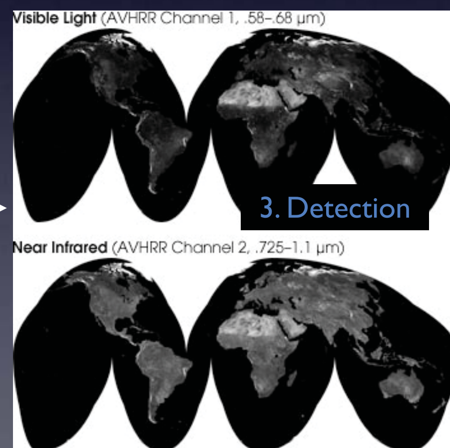
1. Surface reflectance

1. *Scattering* (reflectance) & *emission* spectrum *of some surface* or slab of atmosphere
2. *Scattering* and *absorption/emission* along the two-way *transit* to satellite
3. *Detection* at the instrument of resulting radiance (window channels)

2. Atmospheric transmission



http://earthobservatory.nasa.gov/Features/RemoteSensing/remote_04.php



3. Detection

http://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_2.php

Inferring physical properties from radiances

- Radiances are the only measurable. Optimize extraction of physical quantities:
  - Knowledge of physics (e.g., atmospheric windows and source spectra; radiative transfer) guides the selection of channels and/or channel difference
  - Non-scattering bands allow inference of brightness (blackbody) temperature
  - Forward (physical) models
  - Empirical fits (know we detect what we want plus other stuff; hope other stuff looks like random noise)

Physical (forward) modeling

- Relies on a good physical model (e.g., radiative transfer)
  - Beginning with a first guess, simulate radiances
  - Find errors in simulated radiances given obs
  - Adjust first guess (*the hard part*)
  - Iterate until errors converge to a small value
- Data assimilation strategies are closely related to these ideas

Empirical / statistical modeling

- If the physics are hard to pin down, can also fit the data with a statistical model, e.g.:
  - Linear, non-linear regression
  - Neural networks
- Best with large datasets with uniform, error-free sampling over stationary statistics

Many observations build a climate record

- Climate: assessment of variability and trends require observational stability
- Sufficient number of samples (longer than the life of one instrument)
- Need to take extra care with quality control
  - e.g., GSICS
    http://www.star.nesdis.noaa.gov/smcd/spb/calibration/icvs/GSICS/index.php
  - Instrument / detector degradation
    - Compare radiances prior to other processing to check detector stability (e.g., scattering from bright, deep convective cloud tops)
  - Validation with or training against in-situ or other proxy observations
    - Compare multiple observations after physical inference to check for consistency from system to system

Datasets

Lots of satellite missions = Lots of datasets

Can't cover everything, so we'll
focus on longest-running datasets
in major categories.

Each dataset has limitations due to instruments
and analysis methods.
Proper use is up to you!

Precipitation

- Prior to 1987, records are based on IR
- After 1987, passive microwave era begins with improved space/time resolution and more physically-based retrievals that penetrate cloud top
- TRMM adds radar measurements in 1997 over the tropics
- Upcoming (~2014): GPM, a global follow-on to TRMM — http://gpm.gsfc.nasa.gov/

Precipitation Datasets

- GPCPv2: 5-day and monthly 2.5°, 1979-- and daily1.0°, 1996--
  - Gauge-corrected multi-satellite
  - http://www.gewex.org/gpcpdata.htm
- CMAP: From CPC. Same coverage, data as GPCP. Differences in processing.
  - http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.cmap.html
- TRMM: Tropics, 1997-present. Passive microwave and radar. Profiles, precip features, monthly estimates, multi-satellite merged products.
  - http://trmm.gsfc.nasa.gov/data_dir/ProductStatus.html
- CMORPH: From CPC. Combines output from algorithms of many individual microwave instruments in isolation (AMSU, SSM/I, AMSR, TMI) and propagates them using IR data for precip feature longevity and shape.
  - http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html
- OLR: Longest heritage, looks at anomalies in outgoing longwave for relative (non-quantitative) precip patterns. Based on AVHRR.
  - HIRS: Algorithmic refinement using same OLR data
    - http://www.atmos.umd.edu/~lee/
  - Contributes to GPCP
  - http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.olr.shtml

Sea surface temperature

- Smith & Reynolds, IR and (2002–) microwave T_b
- 1981– ; monthly, 1° and daily, 1/4°
  - AVHRR, AMSR, Ship and Buoy measurements
  - Optimum interpolation method
  - http://www.ncdc.noaa.gov/oa/climate/research/sst/sst.php
- Extended range: 1854 –, monthly, 2°
  - Version 3b includes satellite data in addition to in-situ measurements
  - http://www.ncdc.noaa.gov/oa/climate/research/sst/ersstv3.php

Radiation / Energy Balance

- Top-of-atmosphere radiation balance
- ERBE (1984–; ERBS, NOAA-9, NOAA-10)
  - http://eosweb.larc.nasa.gov/PRODOCS/erbe/table_erbe.html
  - http://iridl.ldeo.columbia.edu/SOURCES/.ERBE/
- CERES (1997– ; TRMM, Terra, Aqua)
  - http://eosweb.larc.nasa.gov/PRODOCS/ceres/table_ceres.html

Near-surface Temperature

- Lower-troposphere record inferred from AMSU brightness temperature
  - Recent studies: Mears and Wentz (2005), http://dx.doi.org/10.1126/science.1114772; Christy et al. (2007), http://dx.doi.org/10.1029/2005JD006881
  - Controversial: Christy claims smaller temperature trend than most studies, with most heating at nighttime.

Land cover

- Many other datasets from UMD / NASA
- Vegetation: NDVI
  - AVHRR (1981–)
    - http://landcover.org/data/gimms/
  - MODIS (2001–)
    - http://landcover.org/data/ndvi/

Snow & Ice

- Visible-based snow extent, 1971–, Northern Hemisphere
  - http://climate.rutgers.edu/snowcover/
- MODIS: 2000–, hi-res daily and monthly cover for snow and ice
  - http://modis-snow-ice.gsfc.nasa.gov/modis.html
- National Snow / Ice Data Center
  - http://nsidc.org

Ozone

- Inferred from solar ultraviolet backscatter
- SBUV instrument
  - http://www.cpc.noaa.gov/products/stratosphere/sbuv2to/sbuv2to_info.shtml

Clouds

- Int'l Satellite Cloud Climatology Project
  - 1983– , global.
  - Also includes snow, ice cover
  - http://isccp.giss.nasa.gov/products/isccpDsets.html

Lightning

- TRMM / LIS, 1997– , -35°–35° latitude
- Microlab OTD 1995–2000,
- Merged LIS/OTD climatology
  - http://thunder.msfc.nasa.gov/data/index.html
- New LIS-only climatology, 0.25°
  - http://cics.umd.edu/~rachela/

More land, ocean, ice

- Many new datasets from the MODIS instrument on Terra and Aqua
  - http://modis.gsfc.nasa.gov/data/
  - http://oceancolor.gsfc.nasa.gov/
- JPL's ocean dataset archive
  - http://podaac.jpl.nasa.gov/index.html

Data wrangling

So far ...

- Many datasets
- Many formats
- Many download sources

- Overwhelming?

Where to start?

- A not uncommon situation: I want to compare diverse datasets, and the software doesn't exist to read and display them both
- I need generic tools that work at a low level to read and plot data
- Does this sound hard to learn?
  - Not really - MATLAB and IDL are essentially these kinds of tools
  - Do the same with free and open tools: Python

A bit of history

- My graduate work: case studies with radar, lightning mapping, and balloon measurements of electric field. Lately, satellite datasets. *So many file formats*, usually at least one of which will be custom.
- Because I work with so many kinds of data, it was hard to identify any one community that produced a domain-specific set of tools that suited me.
- I picked up Python because it's a general and flexible programming language that also has excellent numerics and plotting packages that target general science applications.
- Downside is a bit more setup hassle putting the bits together. For those of you considering a research career, learning to build and install open source tools is a wise investment of time.
- I can help post-course, and/or there are several active and helpful mailing lists for Python and science.

Python tools for science

- NumPy: array-based numerics
  - http://www.scipy.org/
  - NumPy: core numerics. Includes a nice ASCII text parser.
  - SciPy: toolkits for signal processing, lin. algebra, etc. Helpful, not essential.
- Matplotlib: MATLAB-like plotting
  - http://matplotlib.sf.net
- NetCDF3: Pupynere (Pure python reader - no need to compile/install NetCDF)
- NetCDF4-Python (Jeff Whittaker of NOAA CDC)
- PyHDF for HDF4
- PyTables for HDF5
- PyDAP: Interface with OpenDAP from Python (no compiling!)
- PyProj: Python wrapper of a trusted USGS map projections library (in C)
- Academic users: get everything you need for free in the Enthought Python Distribution: http://enthought.com/products/epd.php

Demo

- Combine two unrelated datasets:
  - SST dataset served by OpenDAP (http://nomads.ncdc.noaa.gov/thredds/dodsC/ersstv3Agg)
  - Malnourishment data in a custom ASCII format (http://sedac.ciesin.columbia.edu/povmap/ds_global.jsp)
- Both were new to me