INTELLIGENCE 2025
to improve life on Earth

TAU-UCI
Nov 13, 2018

Efi Foufoula-Georgiou
University of California Irvine (UCI)
THE STATE OF PLAY IN SPACE TODAY

4,000+ satellites orbiting the earth

1,381 active satellites

26% earth observation

52% global communication

10% technology demonstration

4% space science

WHAT IS THE COST OF EARTH OBSERVATION

WHO DOES WHAT IN EARTH OBSERVATION

57% government
26% military
6% commercial
4% civilian

+7% shared systems

Understanding Earth from Space

Courtesy of Michael Freilich, NASA
Earth Science Missions
FY17 Program of Record

ISS Instruments
CLARREO-PF (2020)

JPSS-2 Instruments

Formulation
Implementation
Primary Ops
Extended Ops

Invests/Cubesats
RAVAN (2016)
IceCube (2017)
MiRaATA (2017)
HARP (2017)
TEMPEST-D (2018)
RainCube (2018*)
CubeRRT (2018*)
CIRiS (2018*)
CIRAS (2018*)
CSIM (2018*)
LMPC (TBD)

*Target date, not yet manifested

Courtesy of Michael Freilich, NASA
NASA’s Water and Energy Cycle Missions

Water Cycle Missions

- ICESat
  - Ice elevation
  - Cloud height

- GRACE
  - Column water-content

- TRMM and GPM
  - Global precipitation

- HYDROS
  - Surface wetness
  - Frozen soil
  (SMAP)

Water and Energy Cycle Missions

- EOS-Aura
  - Atmospheric humidity
  - Clouds

- EOS-Terra
  - Snow and ice
  - Vegetation

- CALIPSO
  - Cloud properties

- CloudSAT
  - Cloud profiler

- EOS-Aqua
  - Atmospheric humidity
  - Water storage
  - Clouds
  - Snow and ice

Energy Cycle Missions

- TOMS
  - Total column ozone

- SORCE
  - Total Irradiance measurements

- SAGE
  - Air quality
  - Climate change

- UARS
  - Carbon management
  - Air quality

Planned (not Approved)
- SWOT (Streamflow)
- SCLP (Snowpack)

Complementary Water and Energy Cycle Missions

- QuikSCAT
  - Sea-surface wind velocity

- EO-1 LANDSAT and NMP EO-1
  - Land cover

- NPOESS
  - Global environmental conditions

- GOES
  - Weather

- Aquarius
  - Global sea surface salinity

Courtesy of Michael Freilich, NASA
Exploding Volume of Climate Data from Space

Overpeck et al, Science, 2011
3 Themes @ UCI

1. Precipitation estimation from space
2. Climate dynamics for prediction
3. Landforms supporting life (rivers, deltas)
1. PRECIPITATION

Water cycle dynamics at global to regional scales
Monitoring extremes (hurricanes, tropical storms)
Improving weather and climate models
From TRMM to GPM

Covering 35S to 35N
Microwave Imager (TMI)
-- 9 channels
-- frequencies 10.7-to-85.5 GHz
-- swath width 878 km
Precipitation radar (PR)
-- single-frequency Ku band (13 GHz)
-- swath width 247 km

Covering 68S to 68N
GPM Microwave Imager (GMI)
-- 13 dual-polarized channels
-- frequencies 10.65-183.3 GHz
-- swath width 885 km
Dual Polarization radar (DPR)
-- dual-frequency Ku & Ka (13 and 35 GHz)
-- swath width 120, 245 km
From TRMM to GPM

Covering 35S to 35N
Microwave Imager (TMI)
-- 9 channels
-- frequencies 10.7-to-85.5 GHz
-- swath width 878 km
Precipitation radar (PR)
-- single-frequency Ku band (13 GHz)
-- swath width 247 km

Covering 68S to 68N
GPM Microwave Imager (GMI)
-- 13 dual-polarized channels
-- frequencies 10.65-183.3 GHz
-- swath width 885 km
Dual Polarization radar (DPR)
-- dual-frequency Ku & Ka (13 and 35 GHz)
-- swath width 120, 245 km
GPM Constellation
Passive Microwave Retrieval: An underdetermined Inverse problem

The direct problem:

\[ TB = f(R_C, E_S) \]

Surface emissivity

Hydrometeors profile

Accurately resolved by numerical (physical) radiative transfer models

The inverse problem:

\[ R_C = g(TB, E_S) \]

Underdetermined
Learn patterns from data for retrieval

**Spectral BT**
13-dim space (each point is a BT vector)
*Manifold of BT*

**Rainfall Profiles**
n-dim space (each point is a Z, surf R vector)
*Manifold of R*

"Manifold learning"

**Database**

**Spectral BT**

**Rainfall Profiles**

**Machine Learning and Regularized Estimation in High Dimensional Spaces**

Passive Microwave Retrieval: an Inverse Problem
ShARP: Locally linear embedding for rainfall retrieval

• Inversion Algorithm based on Regularization:
  – Concept of the locally linear embedding (supervised manifold learning):
    - Search for the \textbf{K-nearest neighbors} to detect raining signatures
    \[
    B_S = [b_1 | \ldots | b_K] \in \mathbb{R}^{n_c \times K}
    \]
    \[
    R_S = [r_1 | \ldots | r_K] \in \mathbb{R}^{n_r \times K}
    \]
  – Estimate the \textbf{representation coefficients} and thus the rainfall profile
    \[
    y_i = \sum_{k=1}^{K} c_k b_k + v_k \quad \rightarrow \quad \hat{x}_i = \sum_{k=1}^{K} c_k r_k
    \]

ShARP: Locally linear embedding for rainfall retrieval

- **Detection step:**
  - K-nearest neighborhood search + a probabilistic voting rule for rain/no-rain

- **Estimation Step:**
  - Estimation of the representation coefficients
    \[
    \min_{\mathbf{c}} \left\| W^{1/2} (\mathbf{y} - \mathbf{B}_S \mathbf{c}) \right\|_2^2 + \lambda_1 \| \mathbf{c} \|_1 + \lambda_2 \| \mathbf{c} \|_2^2
    \]
    \[
    \text{subject to } \mathbf{c} \succeq 0, \quad 1^T \mathbf{c} = 1,
    \]
    \[
    \ell_p \text{-norm: } \| \mathbf{c} \|_p^p = \sum_i |c_i|^p
    \]
    \[
    \lambda_1, \lambda_2 > 0
    \]
    \[
    \mathbf{B}_S = [\mathbf{b}_1 | \ldots | \mathbf{b}_{i-1} | \mathbf{b}_i | \mathbf{b}_{j-1} | \mathbf{b}_j | \ldots | \mathbf{b}_K] \in \mathbb{R}^{n_c \times K}
    \]
  - L1-L2 regularization for stability and reduced estimation error
  - Rainfall estimates
    \[
    \hat{\mathbf{x}} = \mathbf{R}_S \hat{\mathbf{c}}
    \]
Yet, lacking performance in several places of the world

Effective Resolution (ER) of NASA’s GPROF v7 (GMI vs KuPR)

- Local values computed from all observations in $3^\circ \times 3^\circ$ boxes.
- March 2014 to February 2017: 16,500 GPM orbits

Learn from the spatial structure in the TB space

1) Sensor Geometry: with different channels responding to different altitude levels, the multi-spectral signature characterizing a given vertical column may be split across several pixels.

2) Specific spatial patterns of TBs are the signatures of specific atmospheric features

QU: HOW TO IDENTIFY THE NEIGHBORHOOD AND HOW TO LEARN FROM IT?
It becomes a very high dimensional problem! Need to learn features!
New Direction: Retrieve patterns not a pixel at a time

Learn from the spatial structure in the TB space

1) Sensor Geometry: with different channels responding to different altitude levels, the multi-spectral signature characterizing a given vertical column may be split across several pixels.

2) Specific spatial patterns of TBs are the signatures of specific atmospheric features

- Lower 37V =>
- Lower emission signal =>
- Lower precipitation?

  **NO!** It is an ice scattering signal =>
  Very active convective Cell
2. SEASONAL PREDICTION

How to best combine climate observations and models to improve predictions?
Bridging the gap: weather and climate

NAS (2016) *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*

36-hr and 72-hr ahead weather forecasts are getting better and better...

*NCEP Operational Forecast Skill*
36 and 72 Hour Forecasts @ 500 MB over North America
[100 *(1-SI/70) Method]

36 hrs 72 hrs

15 Years
Bridging the gap: weather and climate

36-hr and 72-hr ahead weather forecasts are getting better and better…

Our prediction skill on subseasonal (week timescales) to seasonal (S2S) timescales is still very limited.

NAS (2016) Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts

Combining physics and statistics

- Climate models show limited skill in predicting seasonal precipitation months ahead
- Best approach to predict is combining our physical understanding with statistical tools:

\[ Y = f(X) + e \]

Regional hydroclimate (e.g. precipitation, temperature in California)

Large-scale climate modes (e.g. ENSO see below)

The Great El Niño in 2016
Learning from Big Data

- What are the best sources of predictability for a specific region? Are they changing?
- ML and Network analysis can extract much relevant information from the data to improve prediction.

Example: Precipitation in southwestern US

- Dry and variable hydroclimate
- New climate mode discovered, different than ENSO

Annual precipitation (mm/yr)

- 100-300
- 300-500
- 500-700
- ~ 900
- ~ 1250

(Mamalakis et al., 2018, Nature Communications)
Learning from Big Data

- This new mechanism has been more dominant in modulating SWUS precipitation in the last 3 to 4 decades:
  - Vector $X$ may not include important/new modes if based only on our prior knowledge
  - Function $f$ is not constant through time

- In our new project (funded by TRIPODS+CLIMATE program (NSF grant DMS-1839336), we use machine learning to address this problem:

$$E[y_t] = \langle x_t, \beta \rangle$$ \textit{linear model}

- Where the relative contribution of each feature in $X$ is represented by $\beta$ and is calculated by minimizing:

$$\hat{\beta} = \text{arg min}_\beta \sum_{t=1}^{N} (y_t - \langle x_t, \beta \rangle)^2 + \lambda_{TV} \sum_{j,k=1}^{p} \frac{1}{2} |\beta_j - \beta_k| + \lambda_1 ||\beta||_1$$

\textit{Fit observations} \hspace{4cm} \textit{Spatial dependence} \hspace{4cm} \textit{Sparsity}

(Mamalakis et al., 2018, Nature Communications)
Learning from Big Data

\[ E[y_t] = \langle x_t, \beta \rangle \]

\[ \hat{\beta} = \arg \min_{\beta} \sum_{t=1}^{N} (y_t - \langle x_t, \beta \rangle)^2 + \lambda_{TV} \sum_{j,k=1}^{p} \Sigma_{\sqrt{1/2}} |\beta_j - \beta_k| + \lambda_1 ||\beta||_1 \]

- Preliminary results are promising. We can explain more than 40% of precipitation variability in the out-of-sample period.

- The patterns of \( \beta \) can be used to verify and reveal new mechanisms in the large-scale climate system.

- Such approaches can also be used for climate model diagnostics. Do climate models capture the observed interrelations and how are these projected to change under climate change?
3. LANDSCAPES

What can they tell us about process and how are they changing?
Remotely-sensed global imagery is paving the way for a Global Geomorphology

The two most critical problems:

Automatic extraction of dynamic objects

Automatic extraction of critical information

Robust extraction of rivers from multispectral imagery is a very nuanced problem that must consider: water level at time of image, exposed point bars, mixed pixels at boundaries, and clouds, shadows, snow cover, etc.

Following object identification (mask generation), robust algorithms must be capable of objectively distilling relevant metrics and insights without excessive manual intervention.
Robust extraction of rivers from multispectral must consider:
water level at time of image, exposed point bars, mixed
pixels at boundaries, and clouds, shadows, snow cover, etc.

Following object identification (mask generation), robust
algorithms must be capable of objectively distilling relevant
metrics and insights without excessive manual intervention.
Tools for large-scale mask analysis (single rivers)

Create channel masks

1. Download Landsat imagery
2. Classify image pixels
3. Select N images
4. Classified image bank for single year (low stage only)
5. Stack N images
6. Composite
7. Mesh creation
8. Reach maps

Bankfull Masks

- Single-thread
- Hydraulically connected

Spacetime mapping

Area, km²

Temporal changes

Spatial changes


UCAYALI RIVER

1,300 km long

Can predict cutoffs!
Tools for large-scale mask analysis (river networks)

**RivGraph:**
a Python toolbox for analysis of deltaic and braided river channel networks

Extracting channel network topology


Coming soon to a Conda Repository near you…

Morphological properties and topologic metrics
Arctic Deltas (ADs)

The Arctic: North of 66° 33’N

- Climate change affects poles with greater intensity i.e. Polar Amplification (Serreze et al. 2009)
- ADs have on the order of 91 39 Pg-Carbon (Schuur et al., 2015)
- Lakes and ponds are significant sources of methane (i.e. further warming) (Wik, 2016)
- ADs are uniquely characterized by strong spring flooding, permafrost presence, and lake abundance (Walker, 1999)
Can we infer subsurface hydrologic connectivity from the observed (surface) topology and connectivity of lakes and channels in ADs?

**Approach:**
We interrogate lake shrinkage rates and show that distance from the delta channel network controls lake shrinkage and thus subsurface connectivity.

![RivGraph for CN extraction](image)
What are we really studying?

INTELLIGENCE 2025
to improve life on Earth

FINDING THE SIGNAL IN THE NOISE
Workshop on
Data Analytics for Climate and Earth (DANCE): Causality, patterns and prediction

March 27-29, 2019
Arrowhead, CA (USA)

efi@uci.edu