Regression based Modeling of Vegetation and Climate Variables for the Amazon Rainforests

Ankush Khandelwal1, Anuradha Kodali1, Marcin Szubert4, Sangram Ganguly4, Joshua Bongard1, Kamalika Das1
AGU Fall Meeting, San Francisco, December 2013

1University of Minnesota, 2UC Santa Cruz/NASA Ames Research Center, 3University of Vermont, 4BAERI/NASA Ames Research Center. Contact: kamalika.das@nasa.gov

Problem Definition

- Develop methodology to correlate data related to ecosystem dynamics, climate factors, anthropogenic disturbances and extreme events.
- Identify relationships among measured parameters: Model as a regression problem
  - Precipitation, temperature
  - Tropical Rainfall Measuring Mission (TRMM)
  - Land Surface Temperature (LST)
  - Vegetation characteristics: Normalized Difference Vegetation Index (NDVI)
  - Drought, heat waves, forest fires, irrigation
- Compare baseline linear Regression methods with symbolic regression-based Genetic Programming (GP)
- Non-linear dependencies explored between predicted variable and regressors
- Demonstrate technology to answer 2 extreme questions:
  - Impact of 2 Amazon droughts, 2005 and 2010?
  - What are factors impacting vegetation anomalies?
  - How do vegetation factors vary globally?

Case Study: Amazon Forests

Long-term decline of the Amazon carbon sink

Temperature: A snapshot of usual forest drought stress and tree mortality

Problem Formulation

- Point-to-point temporal regression analysis
- Estimate spatio-temporal dependency of forest ecosystems on climate variables

\[ V = \beta_0 + \beta_1 \text{LST}_i + \beta_2 \text{NDVI}_i + \beta_3 \text{NDVI}_{i-1} + \beta_4 \text{NDVI}_{i-2} + \beta_5 \text{NDVI}_{i-3} + \beta_6 \text{NDVI}_{i-4} + \epsilon \]

Vegetation, \( V \)
\( \text{LST}_i \)
\( \text{NDVI}_i \)
\( \text{NDVI}_{i-1} \)
\( \text{NDVI}_{i-2} \)
\( \text{NDVI}_{i-3} \)
\( \text{NDVI}_{i-4} \)
\( \epsilon \)

ij: pixel location index
i: time index
j: spatial neighborhood of index \( i \)
- \( k \): temporal dependency

Open challenge: 1. Estimating function 
2. Estimating best choices for \( k, ab\)

Data Pipeline

NDVI Anomalies across Years

2003 2004 2005 2006

Trend Analysis: Tile Level Mean of NDVI

NDVI log

LST

TRMM

Predict NDVI

Calculate anomaly

Drought years 2005 and 2010 should have higher anomalies indicating drought years

Testing Results for Years 2004–2010

Drought years 2005 and 2010 have high precipitation

Regression Methods

- Baseline Methods:
  - Least square regression with \( l_2 \) penalty (Ridge Regression)
  - Least square regression with \( l_1 \) penalty (LASSO)
- Support vector regression
- Proposed Method:
  - Genetic programming based symbolic regression: non-linear dependencies

Validation Results on Year 2003

- NDVI Season 2 (mean of previous 2 years) has more prominence to predict
- GP has comparable performance to baseline methods
- GP has more prominence to predict the current NDVI Season 2
- Regressed equation has current LST prominent in the lower right hand side tiles

Testing Trends are high at 2005 and 2010 indicating drought years

Important Regressors (Ridge Regression)

- Regressor NDVI Season 2 (mean of previous 2 years) has more prominence to predict the current NDVI Season 2
- Regressed equation has current LST prominent in the lower right hand side tiles
- Similar behavior is observed with other baseline methods whereas non-linear equation is extracted with GP

Summary and Future Work

- Summary
  - Regressor NDVI Season 2 (mean of previous 2 years) is prominent across all tiles
  - GP is used to extract non-linear dependencies between the predicted variable and regressors
  - GP has comparable performance to baseline methods
- Future work
  - Experiment with combinations of temporal look back and/or spatial effects
  - Perform regression at monthly level instead of seasonal
  - Introduce additional regressors (radiation, forest fire maps, deforestation maps)

Acknowledgement: This research is supported by funding from NASA ESTO-ART Program, under grant number NNX15AH44G.