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

Ankush Khandelwal+, Anuradha Kodali§, Marcin Szubert$, Sangram Ganguly¢, Joshua Bongard$, Kamalika Das§

AGU Fall Meeting, San Francisco, December 2015

+University of Minnesota, §UC Santa Cruz/NASA Ames Research Center, $University of Vermont, ¢BAERI/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 *3 science 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

Problem Formulation

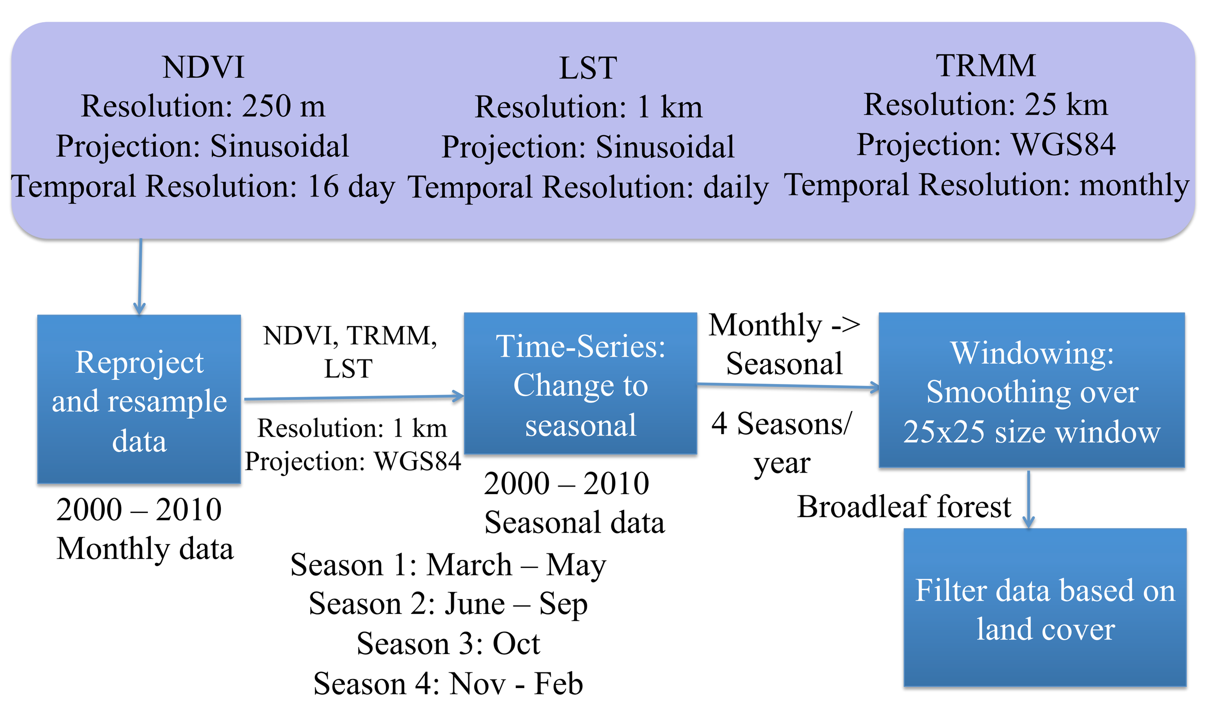
- Point-to-point regression analysis
- Estimate spatio-temporal dependency of forest ecosystems on climate variables

$$V_{ij}^{t} = f(Lc_{ij}, CV_{ij}^{t}, CV_{nb}^{t}, CV_{ij}^{t-1}, CV_{nb}^{t-1}, \ldots CV_{ij}^{t-k}, CV_{nb}^{t-k})$$

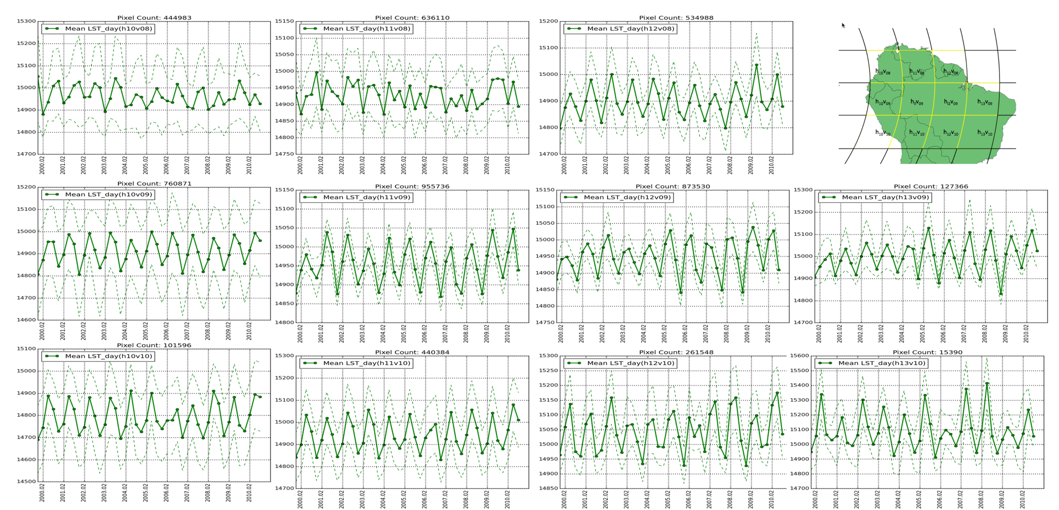
V:vegetation, i,j: pixel location indices
LC:landcover type, t: time index
CV:climate variable(s) nb: spatial neighborhood of index i,j
k: temporal dependency

Open challenges: 1. Estimating function *f*
2. Estimating best choices for *k*, *nb*

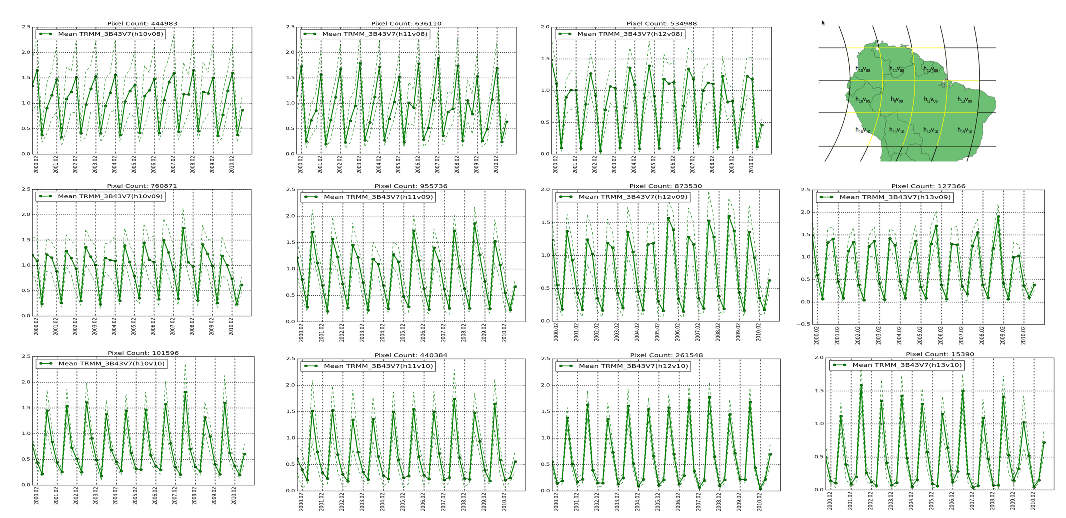
Data Pipeline



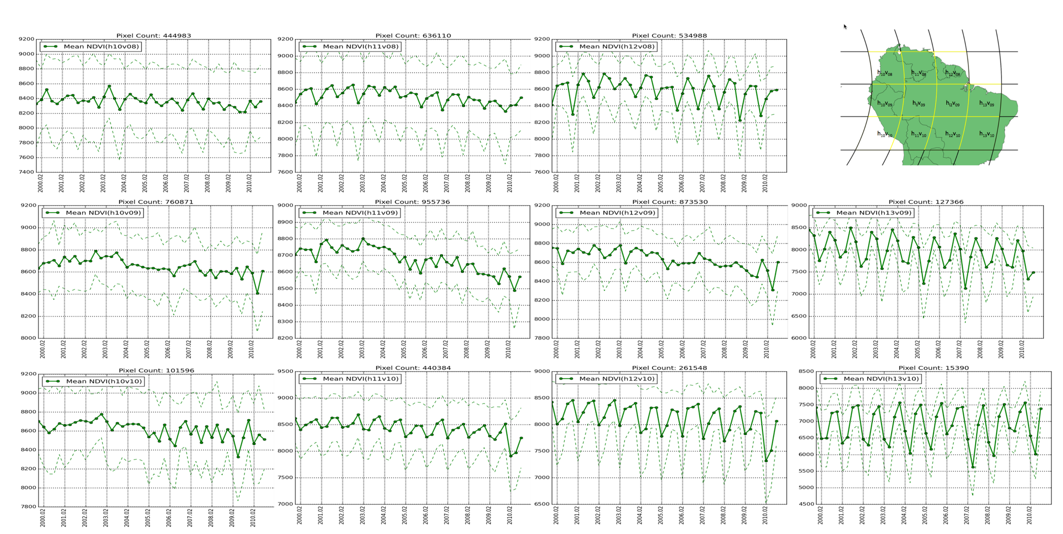
Trend Analysis: Tile Level Mean of LST



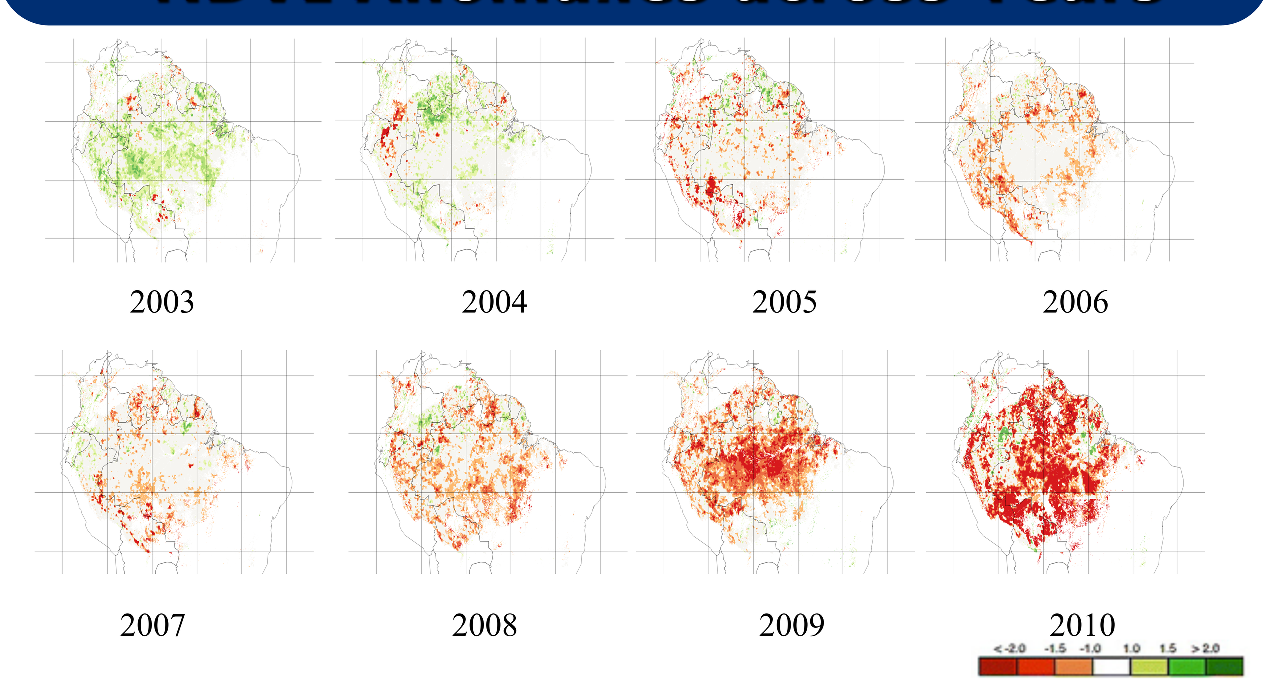
Trend Analysis: Tile Level Mean of TRMM



Trend Analysis: Tile Level Mean of NDVI

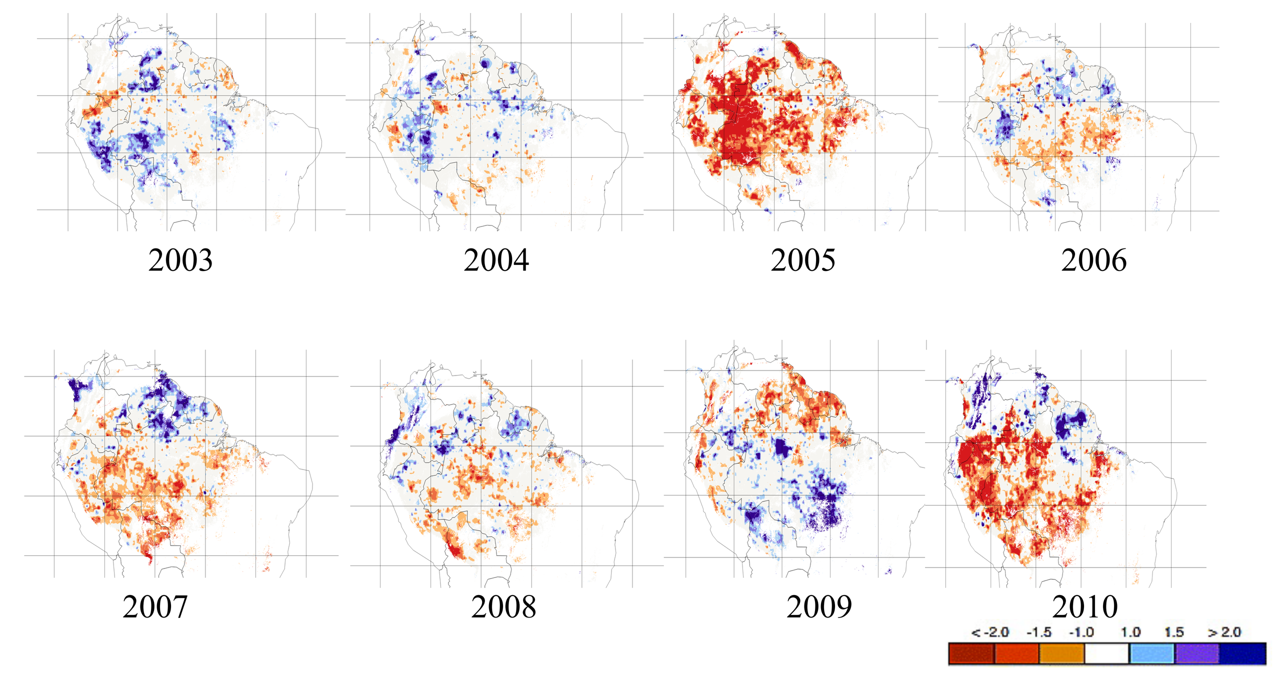


NDVI Anomalies across Years



- Excessive browning in 2010 indicates drought effect

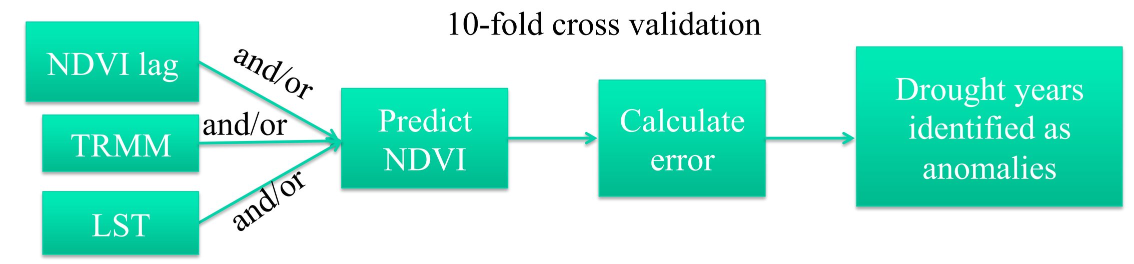
TRMM Anomalies across Years



- Drought years 2005 and 2010 have low precipitation

Problem Set-Up

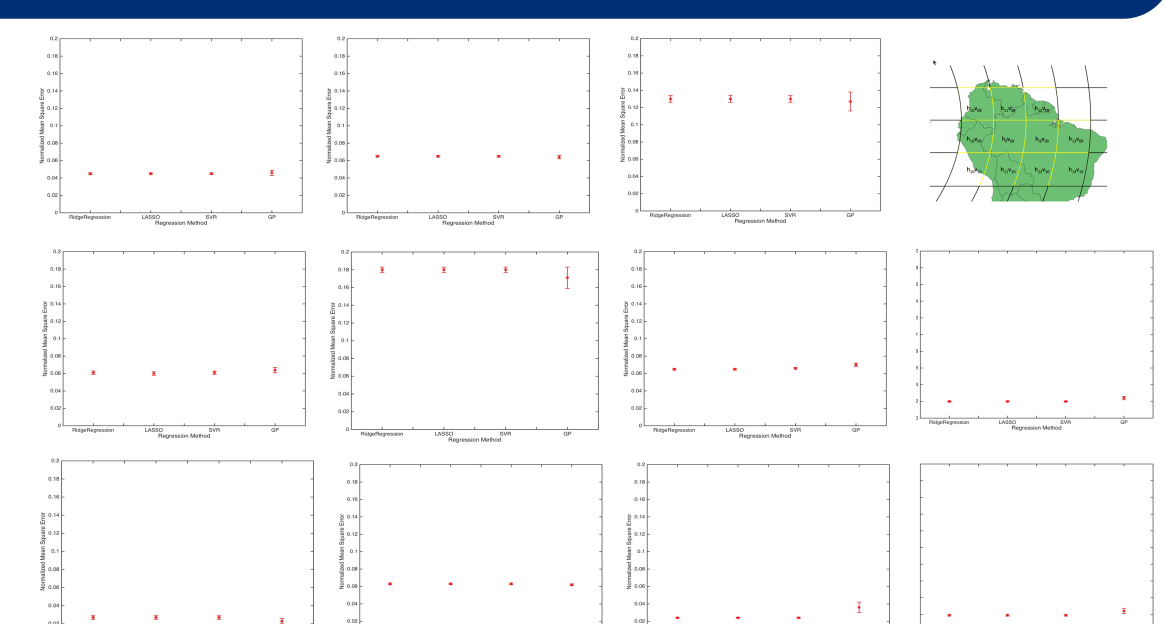
- Modeling scenario: Seasonal mean
  - Dependent variable (target): NDVI Season 2
  - Independent variables (regressors): Historical and current LST and TRMM, historical NDVI
  - Look back: 4 seasons averaged over previous 2 years
  - Total number of regressors: 14
- Metric: Normalized mean squared error
- Parameter selection: 10 fold cross validation
  - Training on 90% data of year 2003
  - Validation on 10% data of year 2003
- Testing on years 2004-2010
- 2005 and 2010 should have higher error indicating drought years



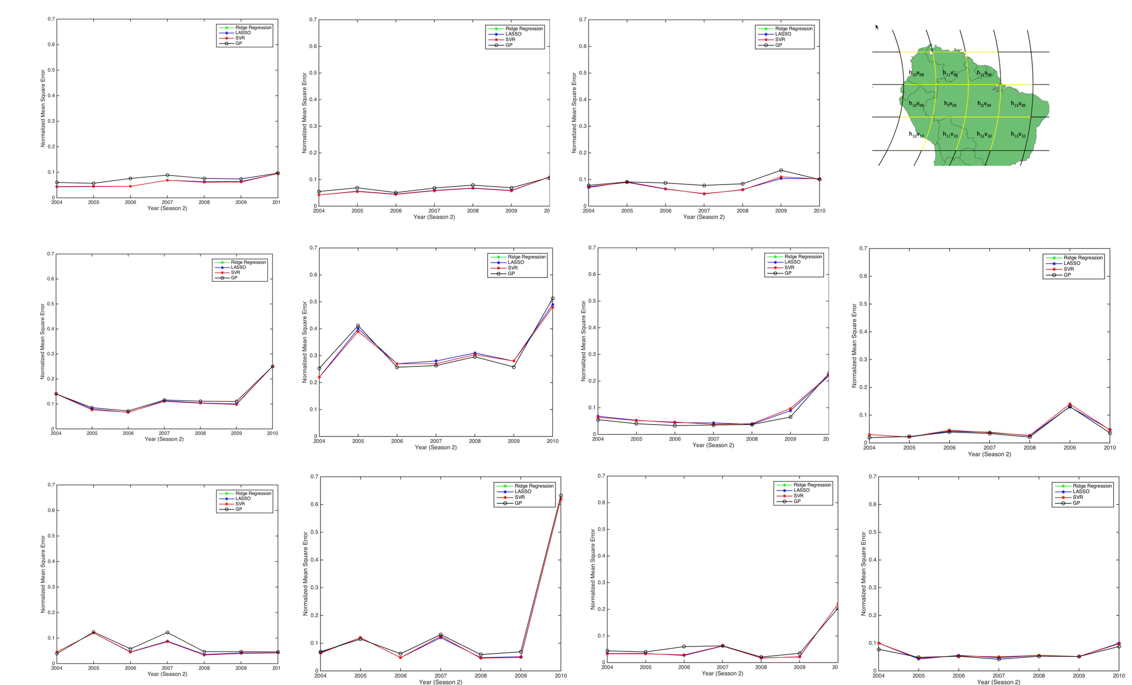
Regression Methods

- 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

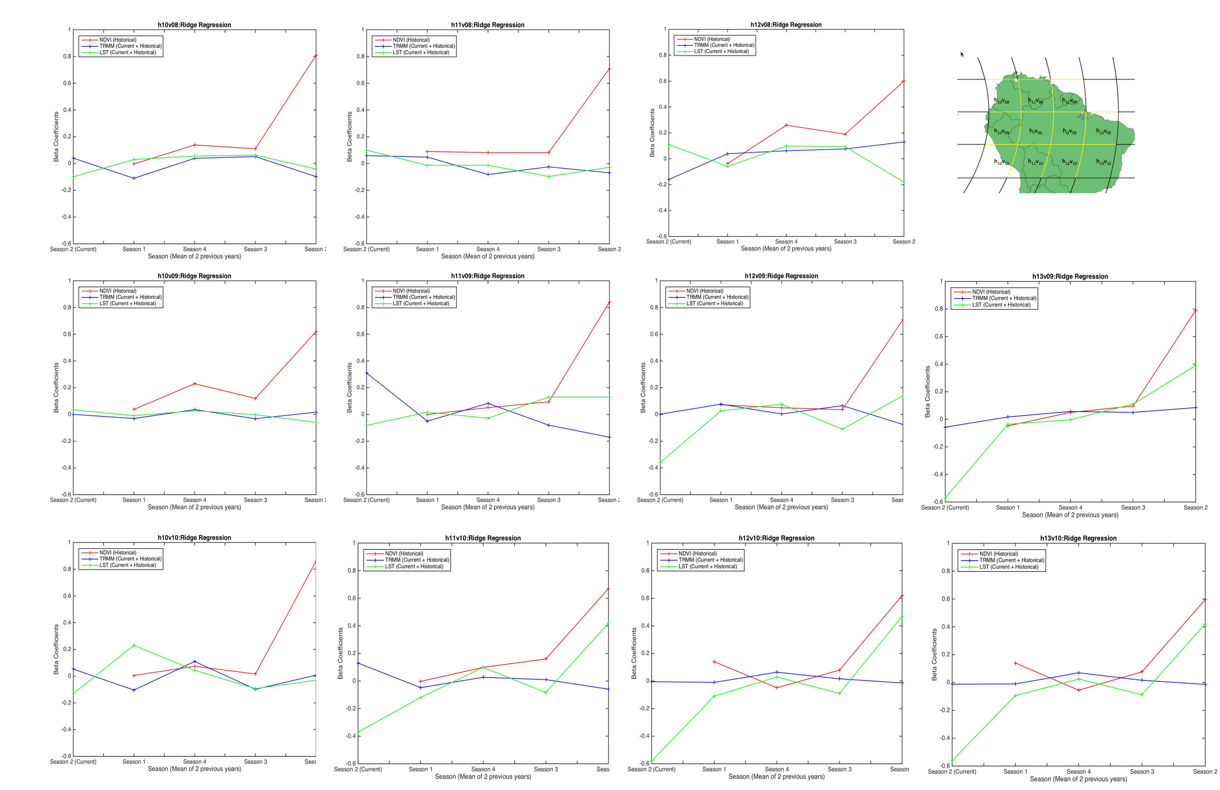


Testing Results for Years 2004-2010



- Testing trends are high at 2005 and 2010 indicating drought years
  - This trend is more prominent in the central region of Amazon

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-AIST Program, under grant number NNX15AH48G