National Aeronautics and Space Administration

Multivariate Fire Prediction

Principal Investigator: Peter Kalmus (398)
Co-Investigators: Hai Nguyen (398), Greg Moore (398), Alphan Altinok (398), Alireza Farahmand (Raytheon)

Program: FY21 R&TD Innovative Spontaneous Concepts

Objective

Create a multivariate fire prediction model using three monthly predictive variables (vapor pressure deficit, VPD; surface soil moisture, SM; enhanced vegetation index, EVI, which measures greenness) with the goal of creating a skillful model with flexibility to easily incorporate new predictors.

Background

As climate change progresses, wildfires are becoming increasingly frequent and severe (Figure 1). Models capable of providing skillful forecasts of fire danger can help guide warning and decision systems. The Fire Danger from Earth Observations (FDEO) model (Farahmand et al. 2020) attempted to predict fire danger over the continental US using monthly VPD from AIRS, SM from GRACE, and EVI from MODIS with a 2-month lead time. It used United States Forest Service (USFS) National Fire Program Analysis Fire-Occurrence Database (FPA FOD) fire data as predictand. The data were collected from 2003-2013 on a monthly 0.25° grid. In each cell, dominant land cover type is determined from US Geologic Survey (USGS) data, and five wildfire regimes (Evergreen, Shrubland, Deciduous, Herbaceous, and Wetland) were analyzed separately. For each land cover type, FDEO performs single-variable quadratic regressions using the variable with highest coefficient of determination. FDEO was not skill-assessed using withheld test data. The FDEO model is not extensible to multiple predictor variables. We seek a systematic and extensible approach.

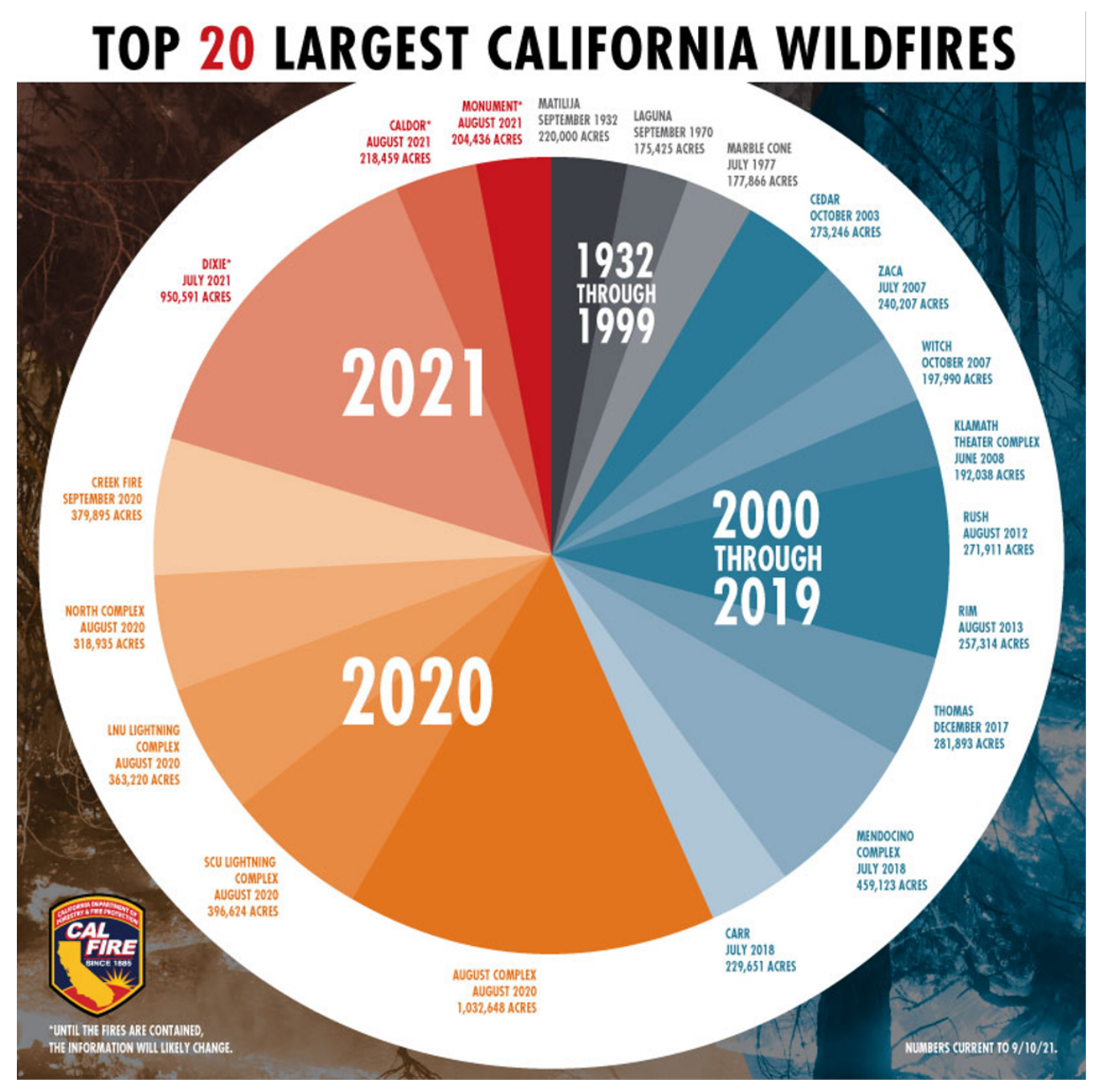

Figure 1. The 20 largest (by acreage) fires in California history demonstrating the very recent emergence of a new climate-change-driven fire regime in California. Figure credit: CalFire.

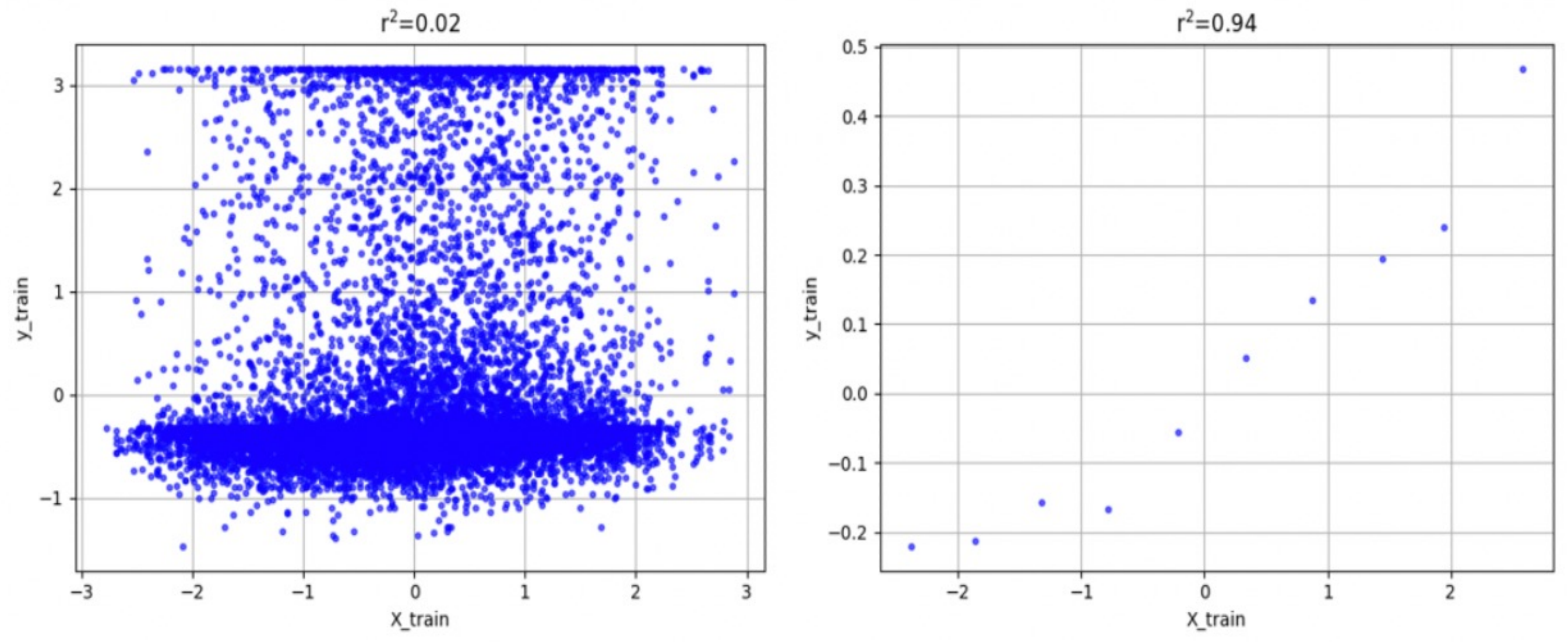

Figure 2: Vapor Pressure Deficit (VPD, 'X_train') standard anomaly unbinned (left) and binned (right) versus fire burned area standard anomaly ('y_train'), for August and the Evergreen lc_type. Figures for Soil Moisture (SM) and Enhanced Vegetation Index (EVI) are similar, i.e. they show correlation with fire burned area when binned but not when unbinned.

Significance/Benefits to JPL and NASA

We implemented a flexible, extensible data and modeling framework for wildfire prediction and compared several models. The project is now primed for follow-on work:

1. Update data through 2021. Western wildfire has entered a quantitatively new regime over the last decade (Figure 5).
2. Include additional predictors, such as daily max, min temperature, daily VPD, daily wind. Weather processes also play a role.
3. Use Bayesian hierarchical methods to perform multivariate modeling with data binning.
4. Focus on key regions, such as the High Sierra in California or the San Gabriel Mountains in California.
5. Provide appropriate uncertainty estimates e.g. from Bayesian models.
6. Explore further model and data transformation options e.g. sigmoid regression.
7. Further validate and assess model interpretability.

Our initial study and this additional work builds JPL's capacity and provides a path toward a multivariate fire prediction model suitable for operationalization. As fires worsen the need for such models as well as funding availability will increase, likely creating opportunities for JPL to contribute in this area for the betterment of society.

Approach and Results

We took the following initial steps:

- Understanding FDEO. We recoded the FDEO model in Python.
- Understanding the data. The three predictor variables show a noisy correlation with the predictand, which only emerges by bin-averaging the predictor and predictand based on the predictor (Figure 2). The predictand is highly imbalanced, with 90% of the raw fire burned area pixels having zero values.
- Creating an object-oriented Python framework for managing and preparing the data. We calculated the climatology, monthly anomaly, and monthly standard anomaly (anomaly divided by standard deviation) for each pixel on a per-land-cover-type and per-month basis. We then performed an N-month running temporal mean, and an M-grid-cell 2D spatial running mean where N and M were nominally set to 3 as per the FDEO model.
- Implementing and calling multiple models from a switch statement using identically-prepared data: a simple quadratic regression (SR) model which mimics the FDEO model, a Random Decision Forest (RDF) regressor, an RDF classifier, a Gradient Boosting Decision Tree (GBDT) regressor, a Bayesian regressor (BR), and a Multi-layer Perceptron (MLP) regressor.
- Assessing the model results. We implemented training/testing splitting using 30% (nominal) withheld data for testing.

We saw no discernible signal when running the models on unbinned data. We therefore:

- Binned the data into bins according to VPD, with 20 bins nominally. We started with a single variable analysis using VPD, and also used a linear combination of the binned VPD, EVI, and SM. This yielded a weak signal.
- Injected synthetic signals into the VPD predictor, and ran the models both on the real data and on the signal-injected data.
- Examined various performance metrics. Figures 2-4 show preliminary results using three of the models (SR, RDF, BR) using August and Evergreen data, using models trained on mean values within K linearly-spaced bins (nominally 20).

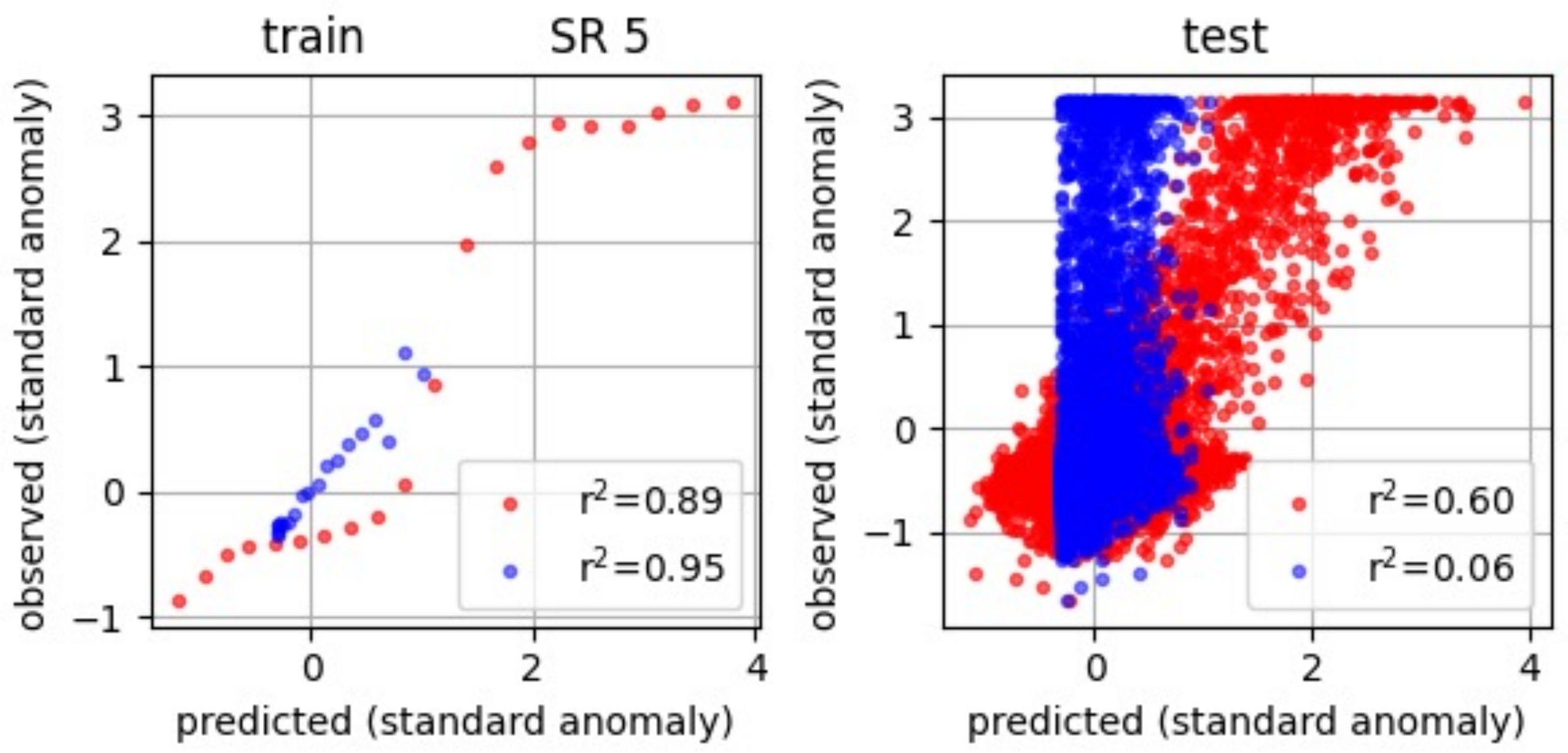

Figure 3: Simple quadratic regression model using VPD only (similar to FDEO). Observed vs. predicted using training and testing data are shown on the left and the right respectively, and data with injected synthetic signal and without are shown in red and blue respectively.

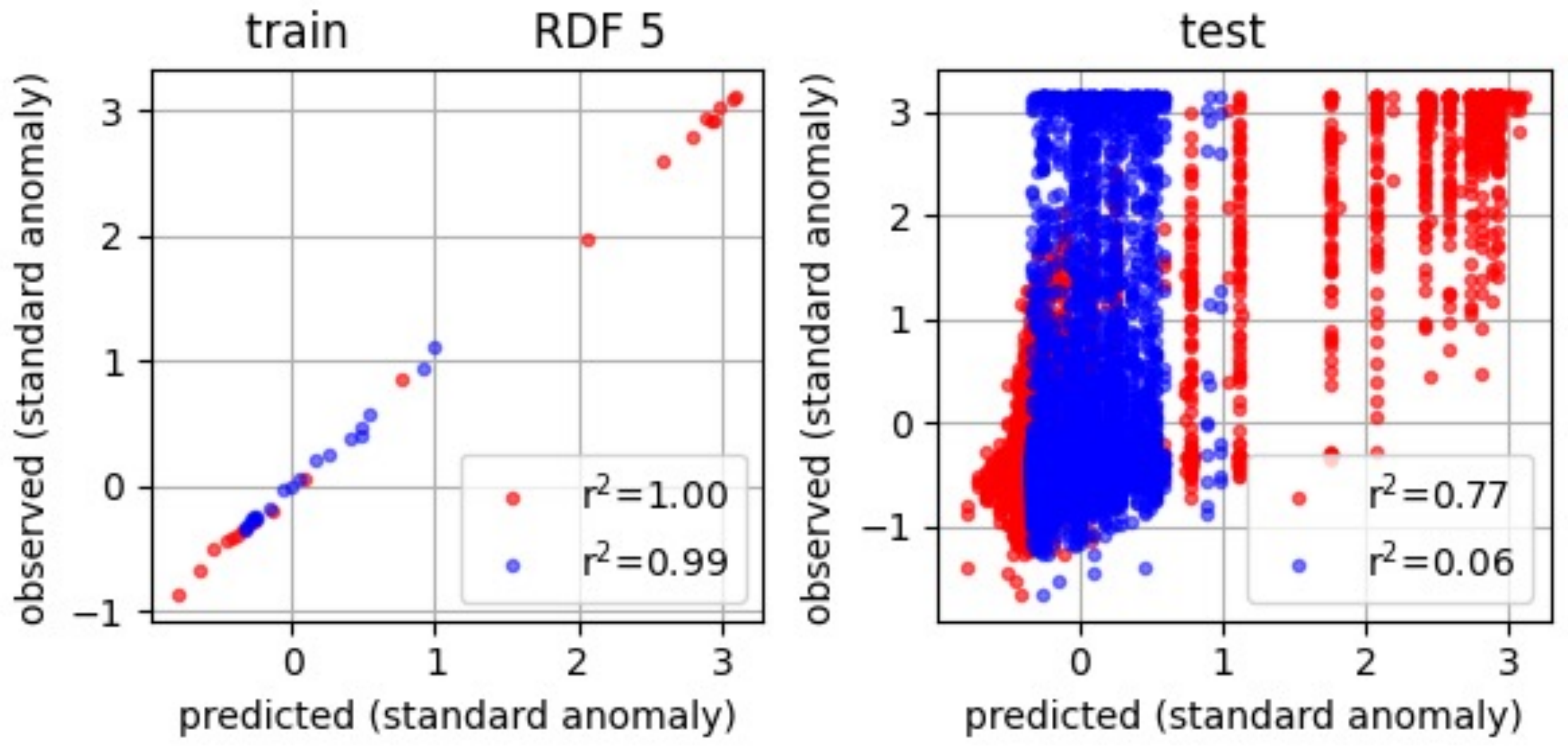

Figure 4: Random Decision Forest (RDF) regression model using VPD, EVI, and SM. Observed vs. predicted using training and testing data are shown on the left and the right respectively, and data with injected synthetic signal and without are shown in red and blue respectively.

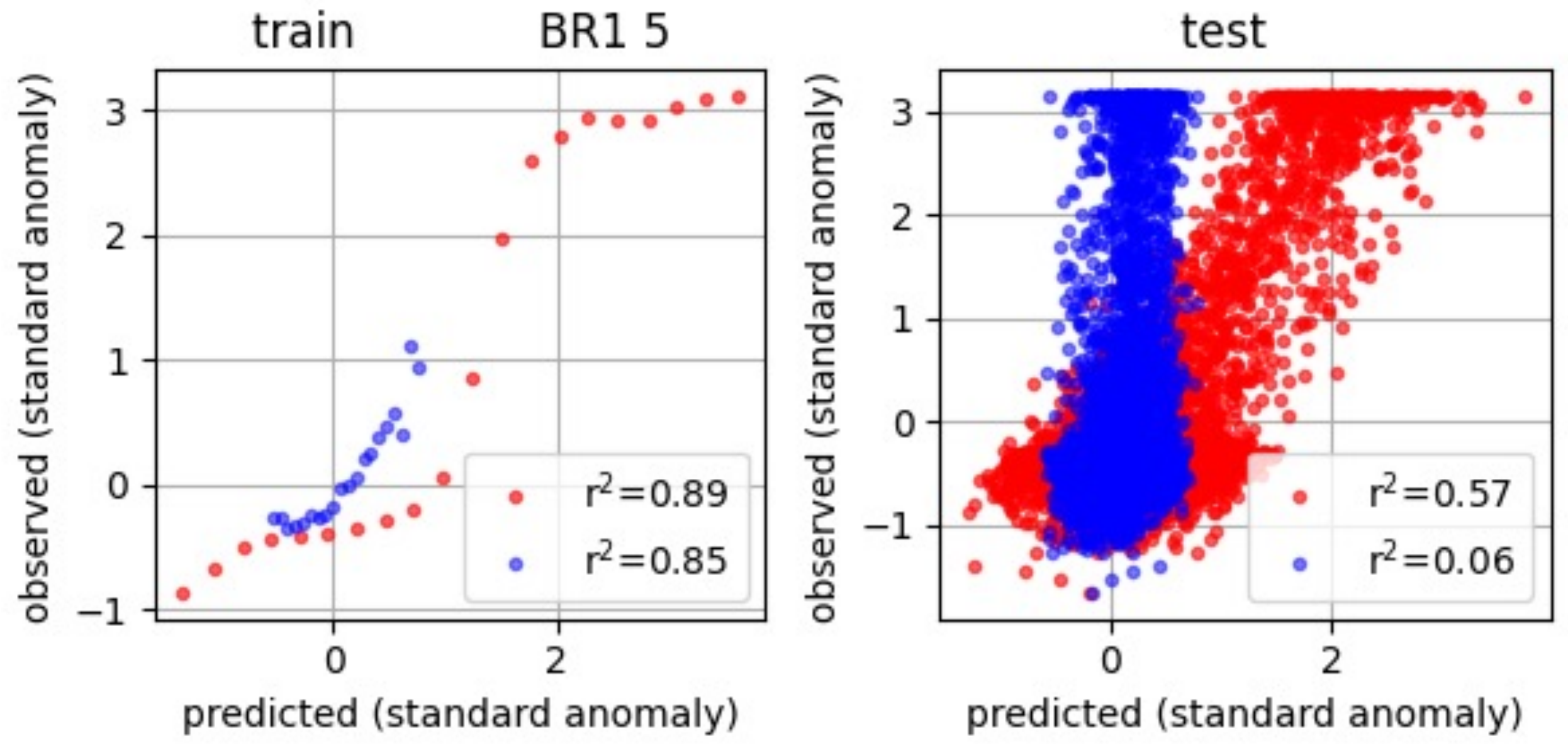

Figure 5: Bayesian regression model using VPD, SM, and EVI. Observed vs. predicted using training and testing data are shown on the left and the right respectively, and data with injected synthetic signal and without are shown in red and blue respectively.

National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California
www.nasa.gov

PI/Task Mgr Contact
Email: Peter.M.Kalmus@jpl.nasa.gov

Clearance Number:
RPC/JPL Task Number: R21252