Predicting Wildfire Fuel Moisture from Atmospheric Data Using Machine Learning

Pankaj K. Jha[1], Jeffrey D. Mirocha[1], Angel Farguell[2], Adam Kochanski[2], and Philp Cameron-Smith[1]

[1]Lawrence Livermore National Laboratory
[2]San Jose State University

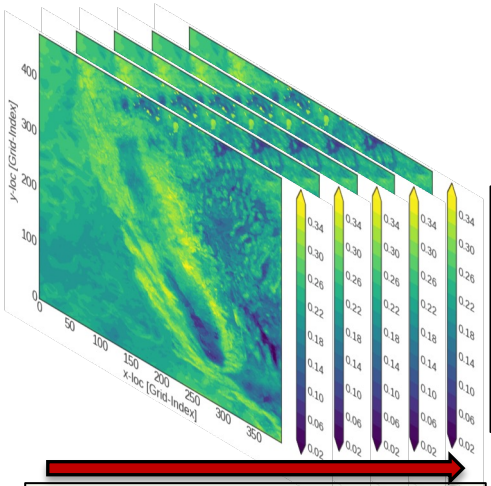
LLNL-PRES-XXXXXX

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 and was supported by the LLNL-LDRD Program under Project 22-SI-008. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

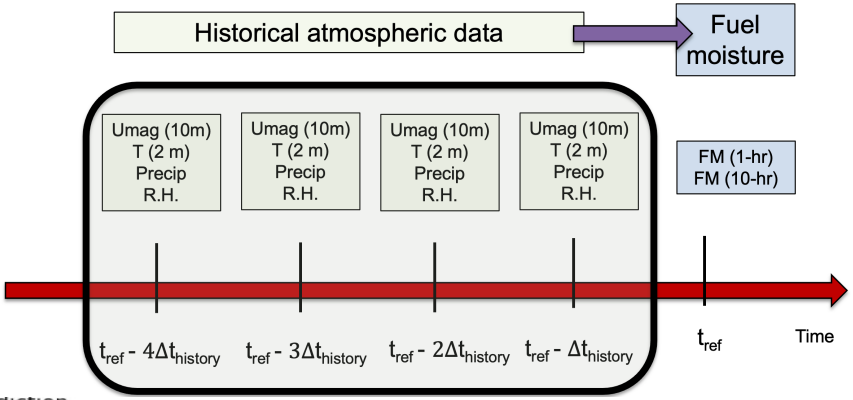
Overview

Machine Learning Automation Framework

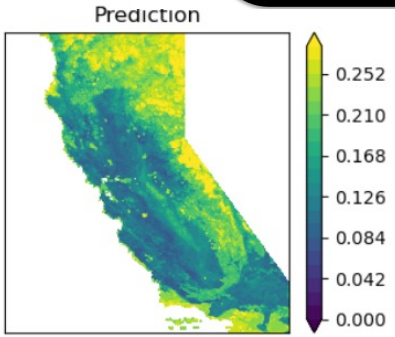


Historical atmospheric and fuel moisture data for 21 years

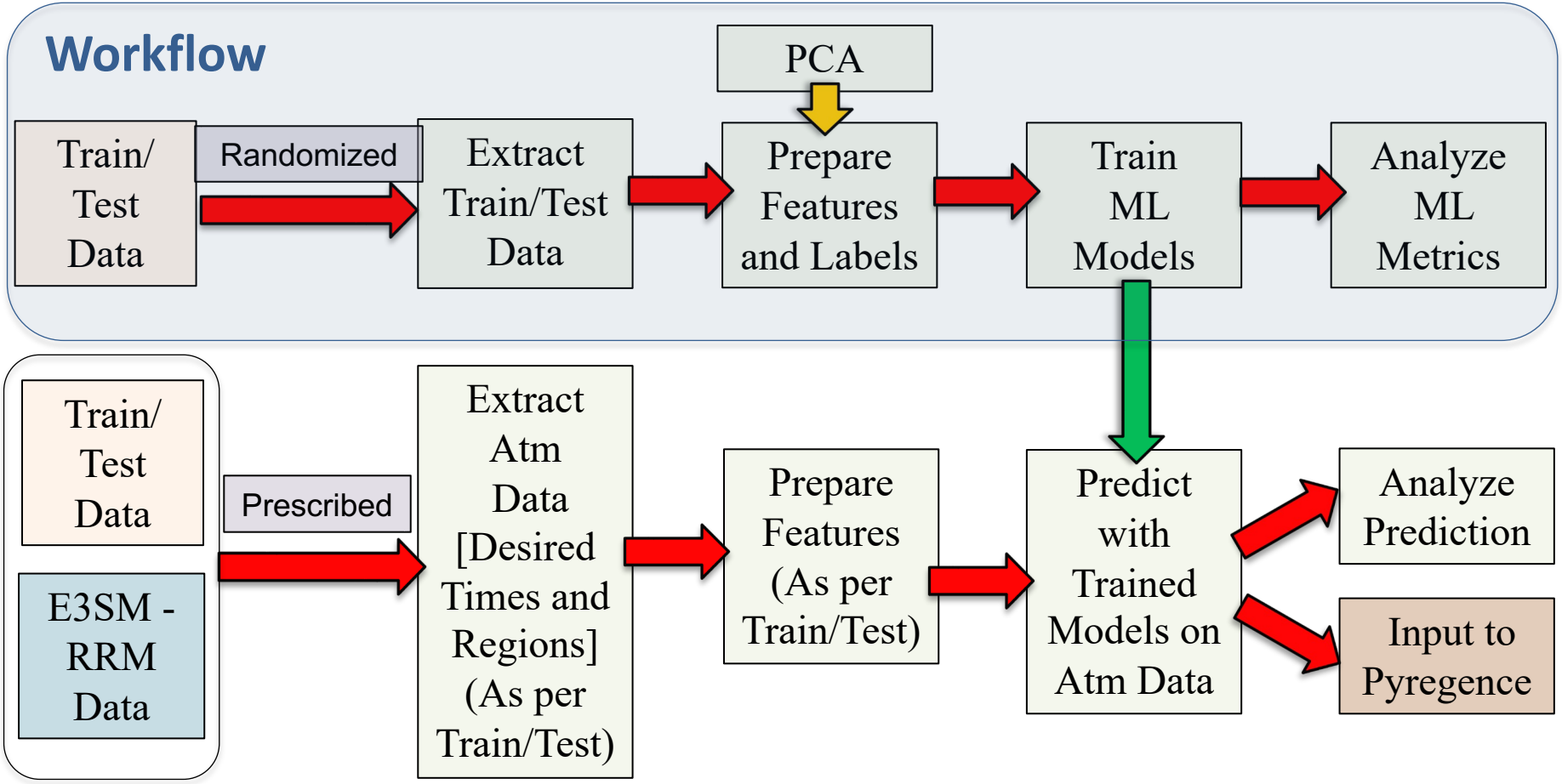
Seeking a relation between fuel moisture (FM) content and preceding atmospheric data



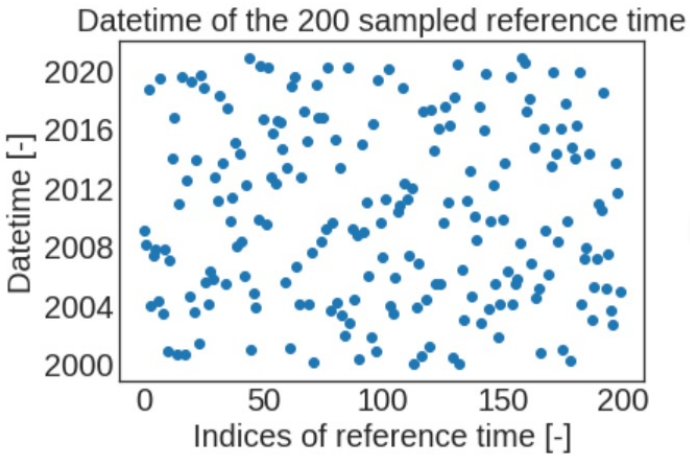
Goal

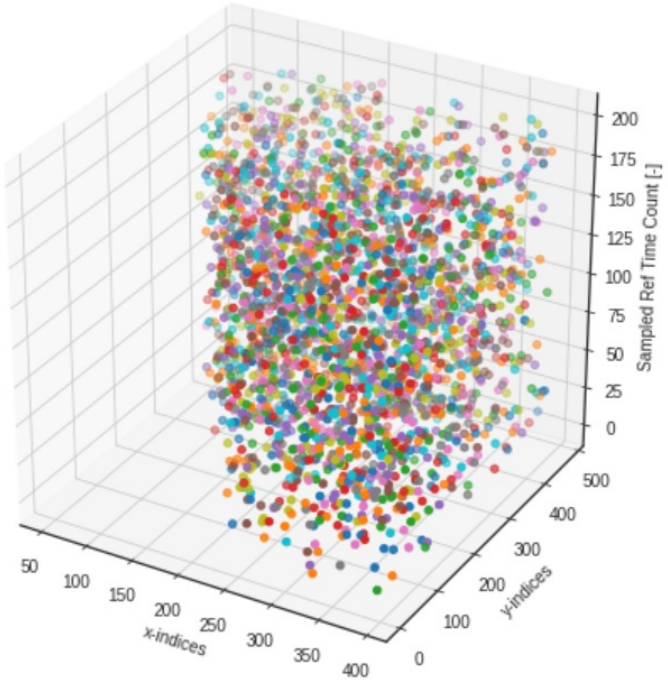


Workflow


PCA
Train/ Test Data
Randomized
Extract Train/Test Data
Prepare Features and Labels
Train ML Models
Analyze ML Metrics
Train/ Test Data
E3SM - RRM Data
Prescribed
Extract Atm Data [Desired Times and Regions] (As per Train/Test)
Prepare Features (As per Train/Test)
Predict with Trained Models on Atm Data
Analyze Prediction
Input to Pyregence

Extract WRF Data





Randomly sample reference time for FM, between 2000-2021

For each reference time, randomly sample grid points

Extract WRF Data

Extracted Raw Data

	FM_time_ind	FM_ts	his_time_ind	grid_ind	j_ind	i_ind	FM_10hr	FM_100hr	HGT	U10[-8hr]	...	T2[-4hr]
998	36601	2004-03-05_01	[36593, 36595, 36597, 36599]	155520	392	288	0.150865	0.102206	1363.334839	3.066367	...	278.576202
999	36601	2004-03-05_01	[36593, 36595, 36597, 36599]	183608	463	260	0.153421	0.136851	1569.290771	4.593234	...	275.457977
1000	113518	2012-12-12_23	[113510, 113512, 113514, 113516]	77031	194	207	0.187562	0.160372	26.139740	0.256508	...	284.830414
1001	113518	2012-12-12_23	[113510, 113512, 113514, 113516]	70744	178	256	0.176889	0.141559	120.549843	-1.779185	...	285.025299

Prepare Training Data

Three types of labels used:
- Actual FM (for regression models in ML)
- Binary
- Multi-class
  - Uniform-interval
  - Non-uniform interval

- Binary label defined by a threshold, i.e., FM > FM_threshold => label ‘0’ or no risk. Otherwise, label ‘1’
- Multi-class (MC) labels assigned for different ranges/levels of FM

	FM_10hr	FM_100hr	FM_10hr_bin	FM_100hr_bin	FM_10hr_MC	FM_100hr_MC
0	0.119865	0.112545	0	0	3	3
1	0.171848	0.123733	0	0	5	4
2	0.230967	0.247327	0	0	7	7
3	0.167400	0.141724	0	0	5	4
4	0.143390	0.147591	0	0	4	4
...	...	...	...	...	...	...
99995	0.057134	0.044610	1	1	1	1
99996	0.075633	0.049422	1	1	2	1
99997	0.077075	0.043984	1	1	2	1
99998	0.156727	0.187614	0	0	5	5
99999	0.026654	0.012576	1	1	0	0

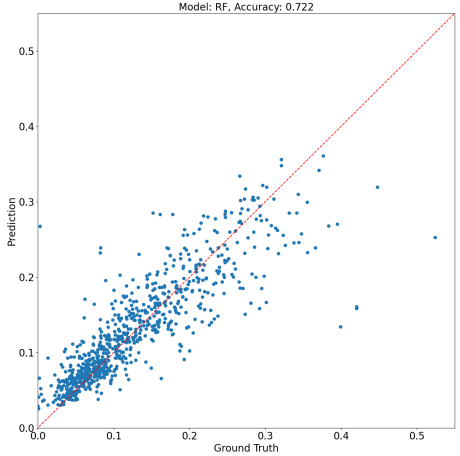
Run Machine Learning (ML) Model

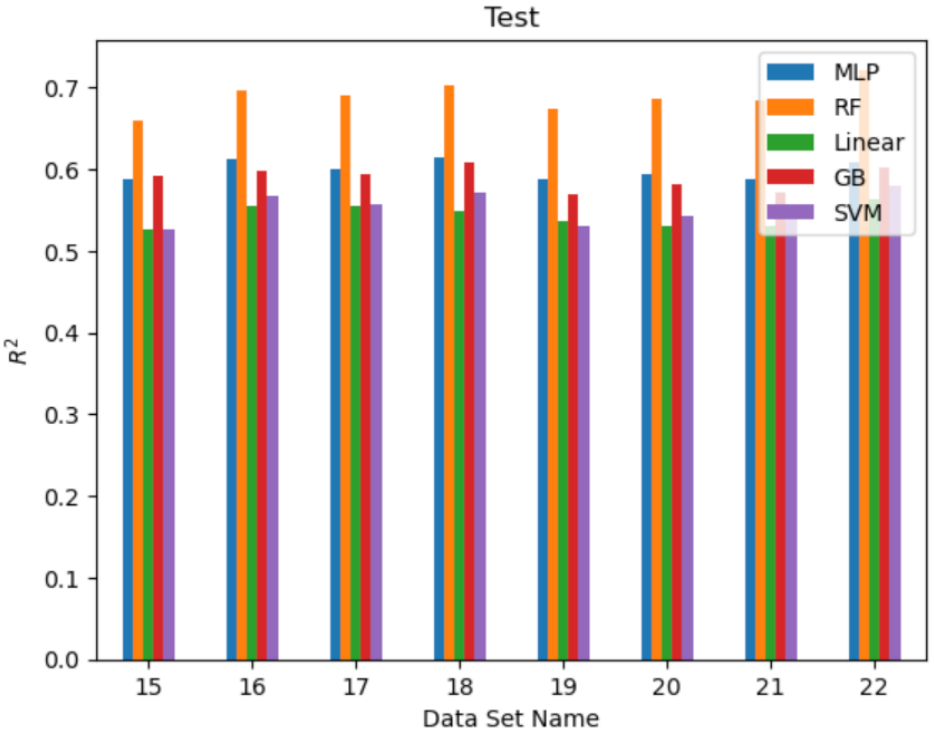
```
def define_model (FM_label_type, model_name):
    if (FM_label_type == 'Regression'):
        match model_name:
            case 'Linear':
                model = LinearRegression()
            case 'SVM':
                model = SVR()
            case 'RF':
                model = RandomForestRegressor()
            case 'MLP':
                model = MLPRegressor()
            case 'GB':
                model = GradientBoostingRegressor()


    elif (FM_label_type == 'Binary' or FM_label_type == 'MultiClass'):
        match model_name:
            case 'SVM':
                model = SVC()
            case 'RF':
                model = RandomForestClassifier()
            case 'MLP':
                model = MLPClassifier()
            case 'GB':
                model = GradientBoostingClassifier
```

Analyze ML Metrics

Regression

	temporal_data [%]	spatial_data [%]	max_history [hrs]	hist_interval [hrs]
15	0.543198	0.595586	16	4
16	0.543198	0.595586	20	4
17	0.543198	0.595586	24	4
18	0.543198	0.595586	28	4
19	0.543198	0.595586	32	8
20	0.543198	0.595586	40	8
21	0.543198	0.595586	48	8
22	0.543198	0.595586	56	8

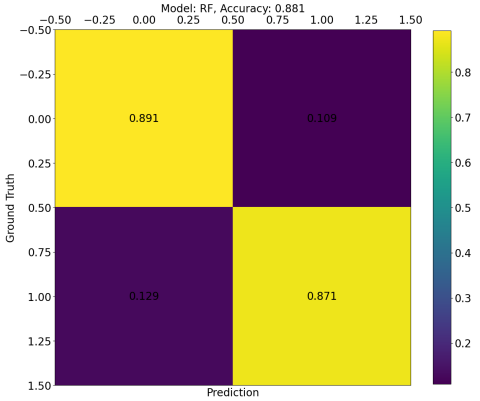


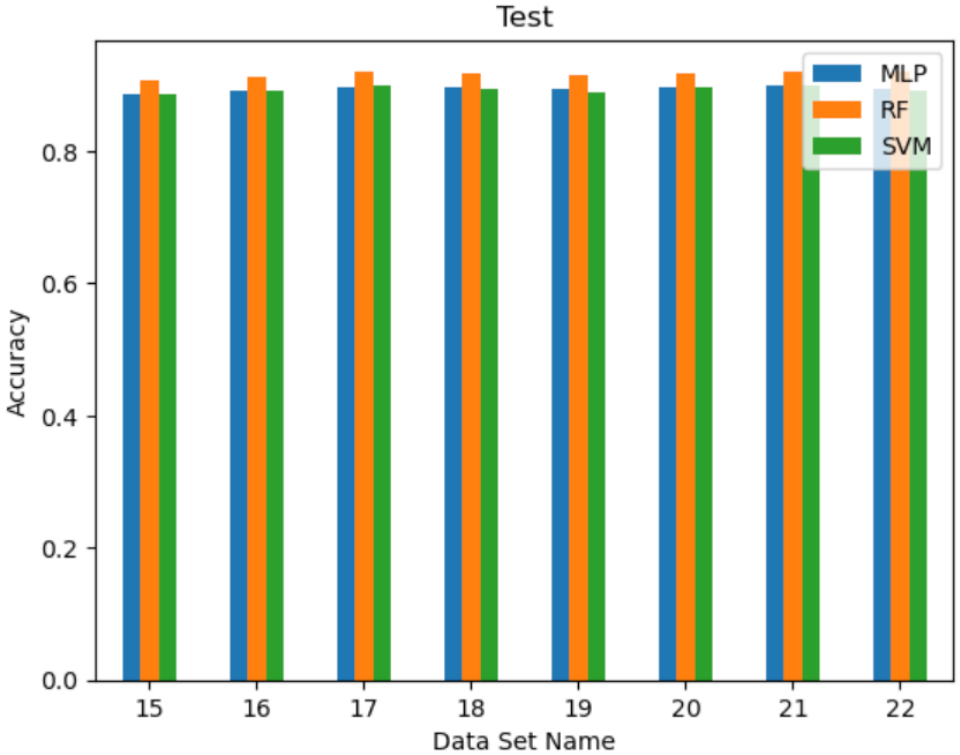


Analyze ML Metrics

Classification

	MLP	RF	SVM
15	0.88722	0.90652	0.88665
16	0.89126	0.91316	0.89239
17	0.89806	0.92023	0.89863
18	0.89778	0.91767	0.89483
19	0.89295	0.91563	0.89016
20	0.89719	0.91702	0.89648
21	0.89874	0.92103	0.89876
22	0.89367	0.91926	0.89261



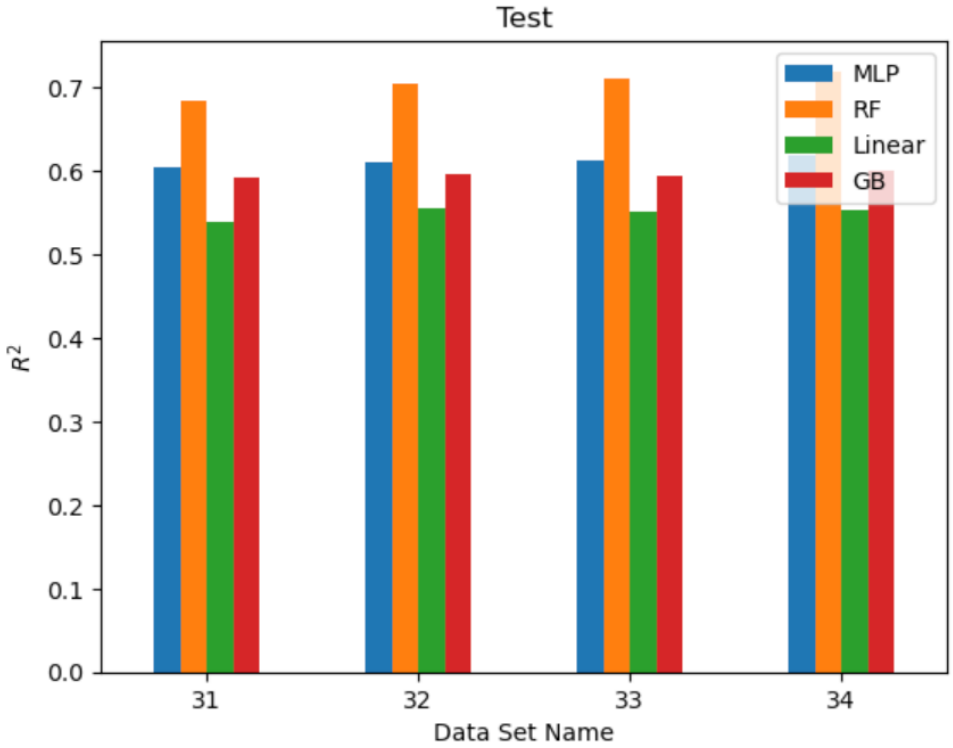


Analyze ML Metrics

Sensitivity Analysis

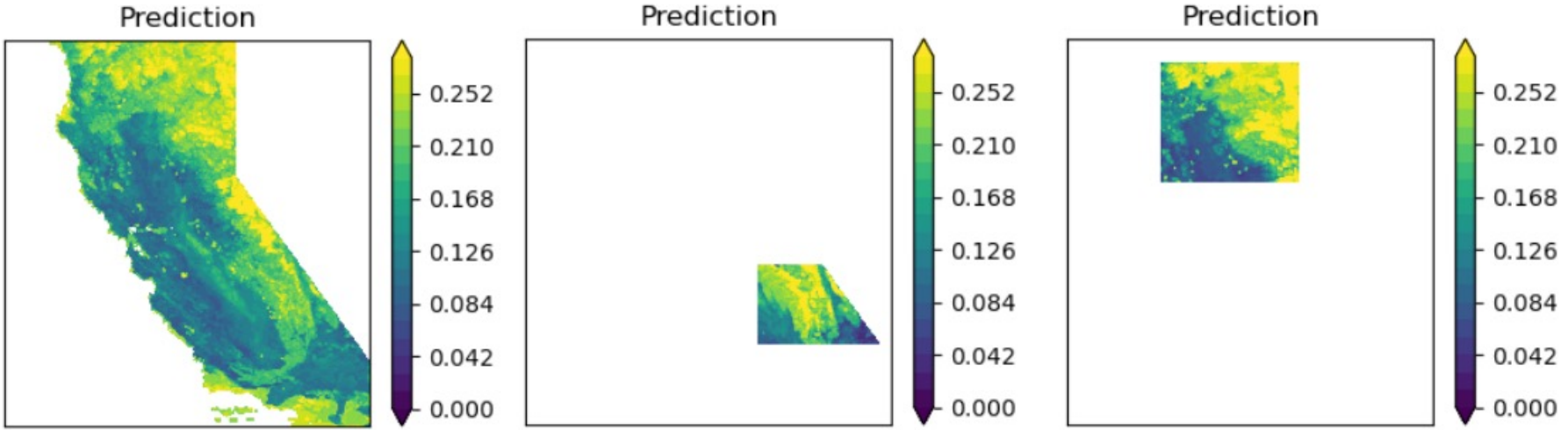
	temporal_data [%]	spatial_data [%]	max_history [hrs]	hist_interval [hrs]
31	1.086396	1.191171	16	4
32	1.086396	1.191171	20	4
33	1.086396	1.191171	24	4
34	1.086396	1.191171	28	4

	MLP	RF	Linear	GB
31	0.605045	0.684777	0.538737	0.591665
32	0.611703	0.704888	0.555720	0.597558
33	0.612999	0.710634	0.551414	0.595053
34	0.619176	0.719822	0.554394	0.601560



Prediction on Prescribed Data

At desired time stamp (2018-11-08_22) and desired region



Future Work

- Study the effect of data sets and model parameters on the performance metrics
  - Identify ways to improve accuracy on randomized test data
  - Analyze predictions at desired times and regions
  - Continue to study the effect of features and model parameters on the accuracy of prediction
  - Study coarser fuel classes (100-hr, 1000-hr)
  - Dimensionality reduction using PCA
  - Apply the best performing model to E3SM-RRM data for future climate assessment
  - Examine transferring predicted FM data into risk probabilities


Lawrence Livermore
National Laboratory