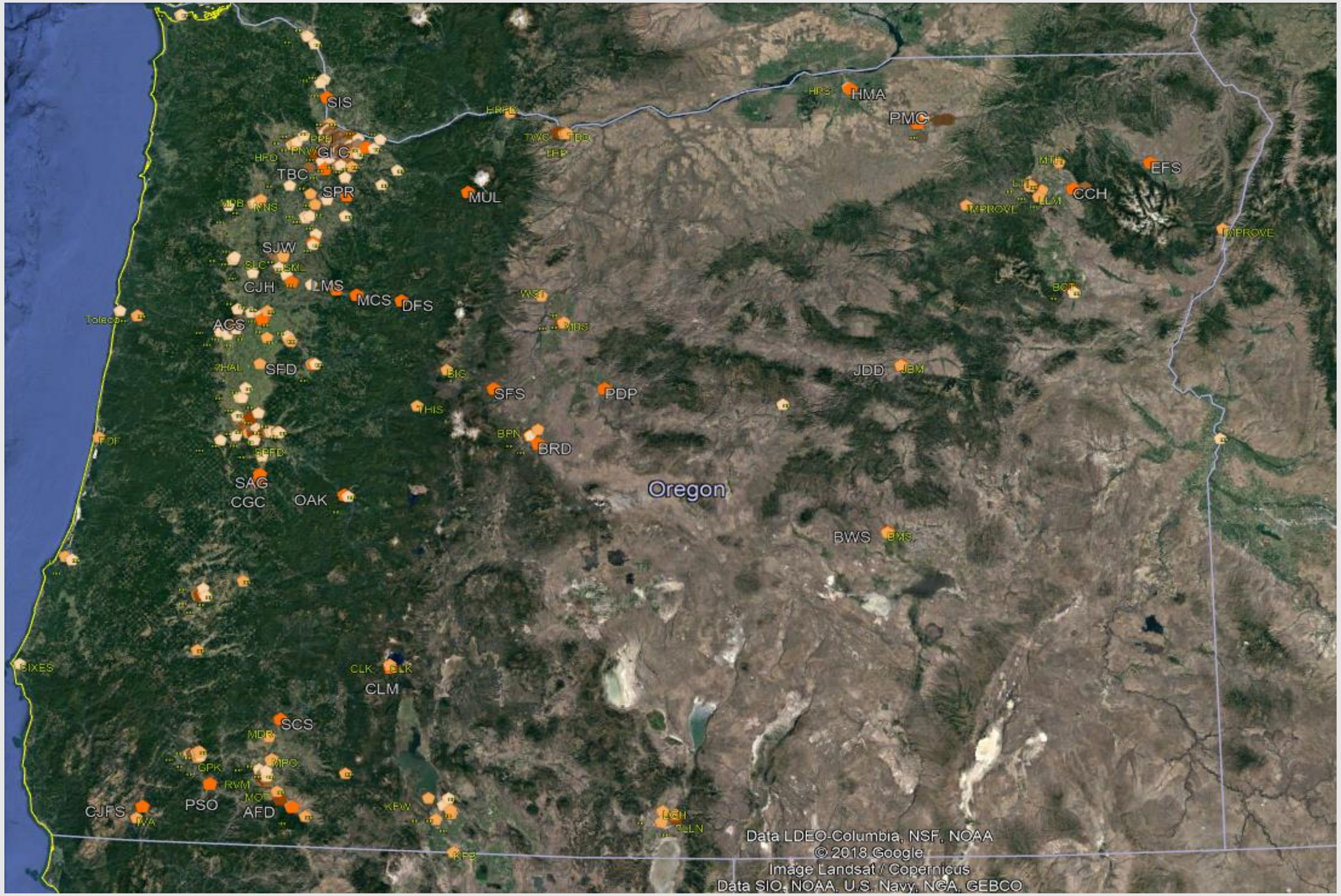
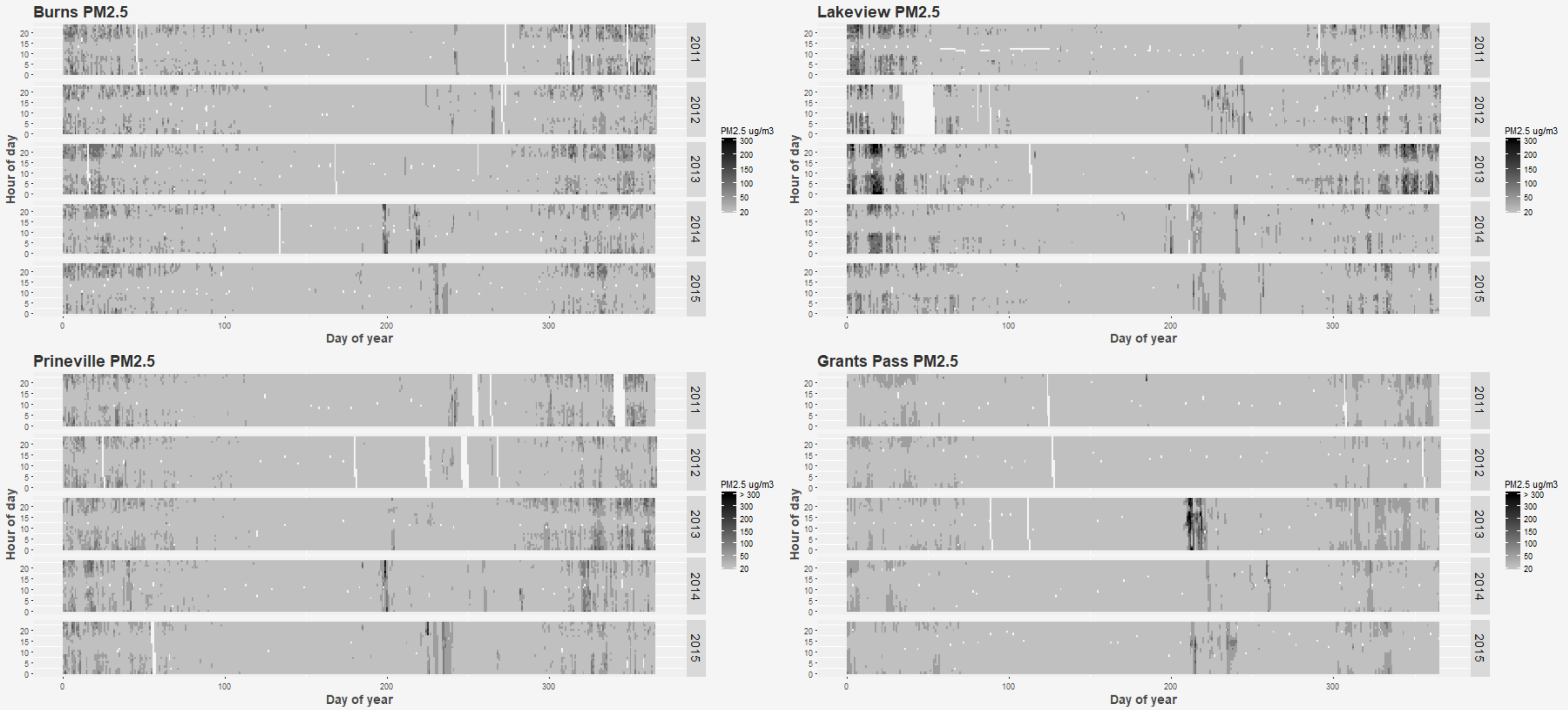
Exploring the Performance of Machine Learning Algorithms in Short-term Forecasting of $PM_{2.5}$

Meenakshi Rao, Lori Pillsbury
Laboratory & Environmental Assessment Division
Oregon Department of Environmental Quality

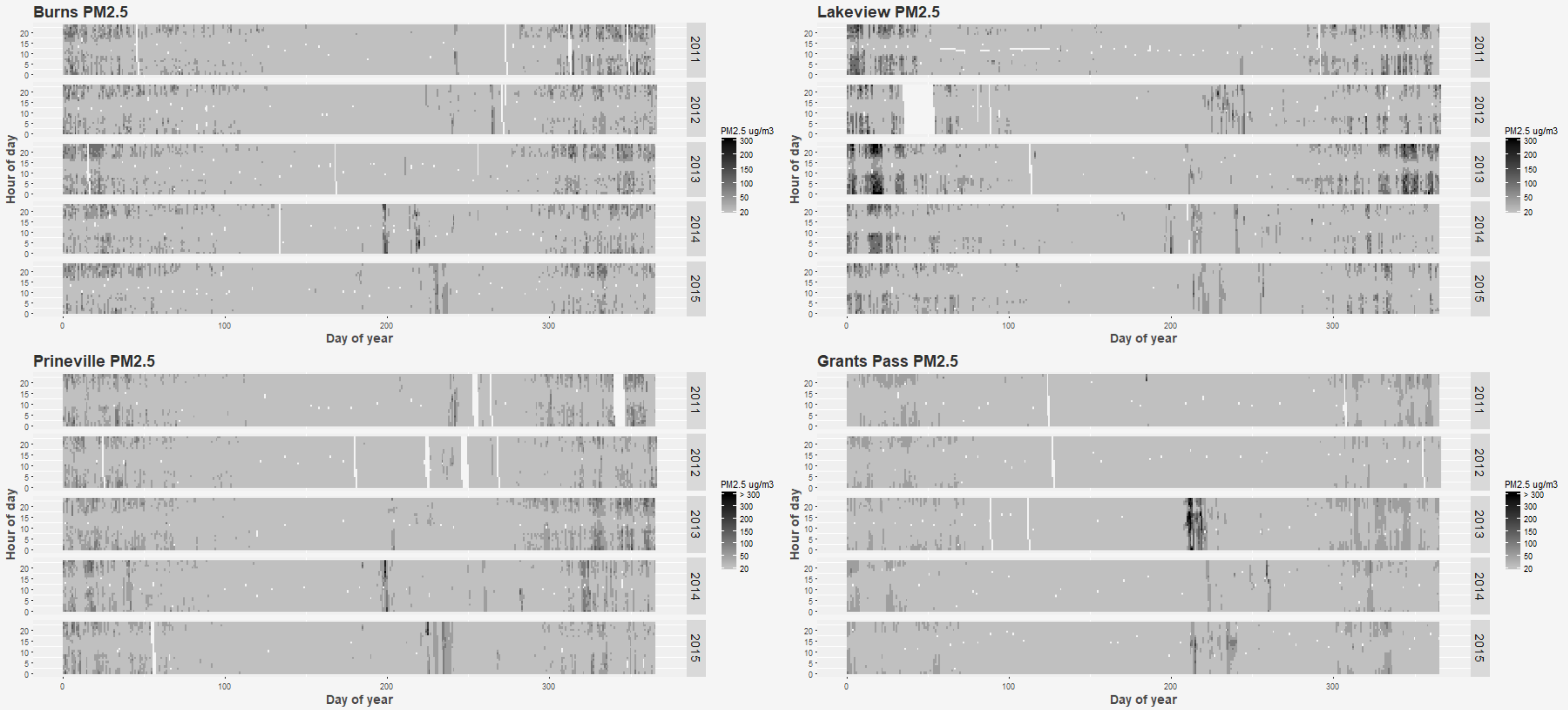


OR DEQ Air Quality Monitoring Network



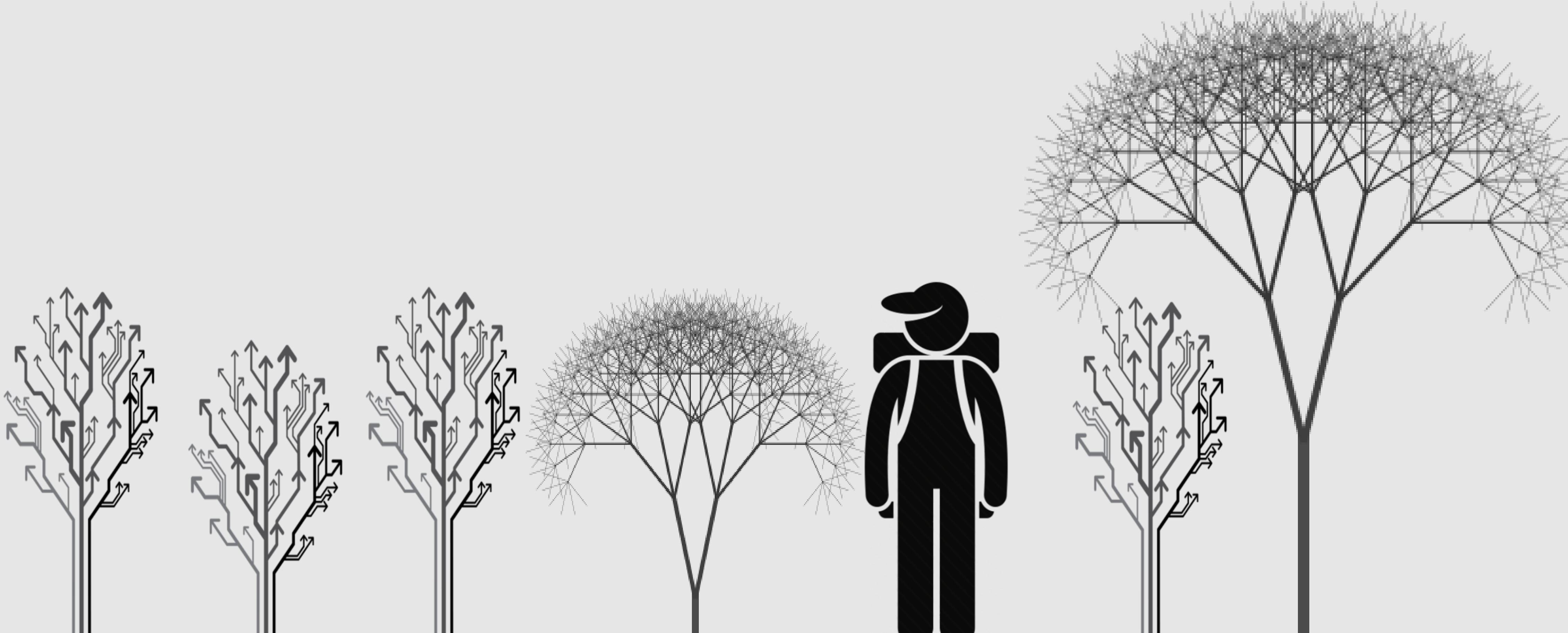
Burns PM2.5
Lakeview PM2.5
Prineville PM2.5
Grants Pass PM2.5
Hour of day
Day of year
PM2.5 ug/m3
2011
2012
2013
2014
2015

A lot of data!



historic trends……. …….future predictions?

Explore Machine Learning!



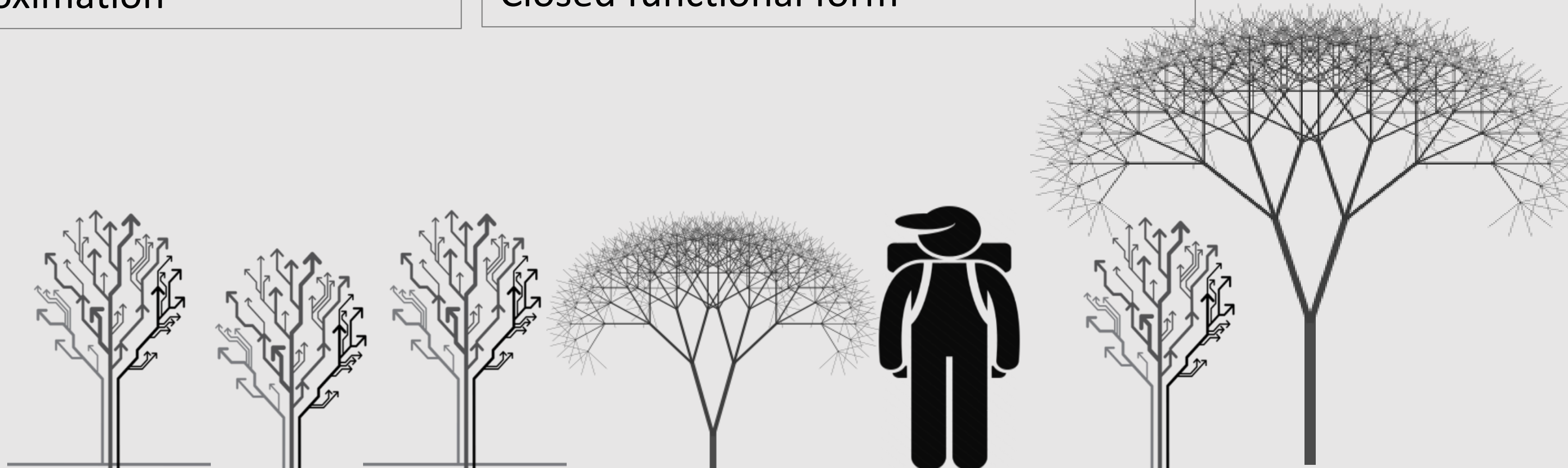
BUT… WHAT IS

MACHINE LEARNING?

MACHINE LEARNING ≡ COMPUTATIONAL STATISTICS

Observations: $x_1, x_2, x_3 \ldots x_n$
Outcome/Result: $y_1, y_2, y_3 \ldots y_n$

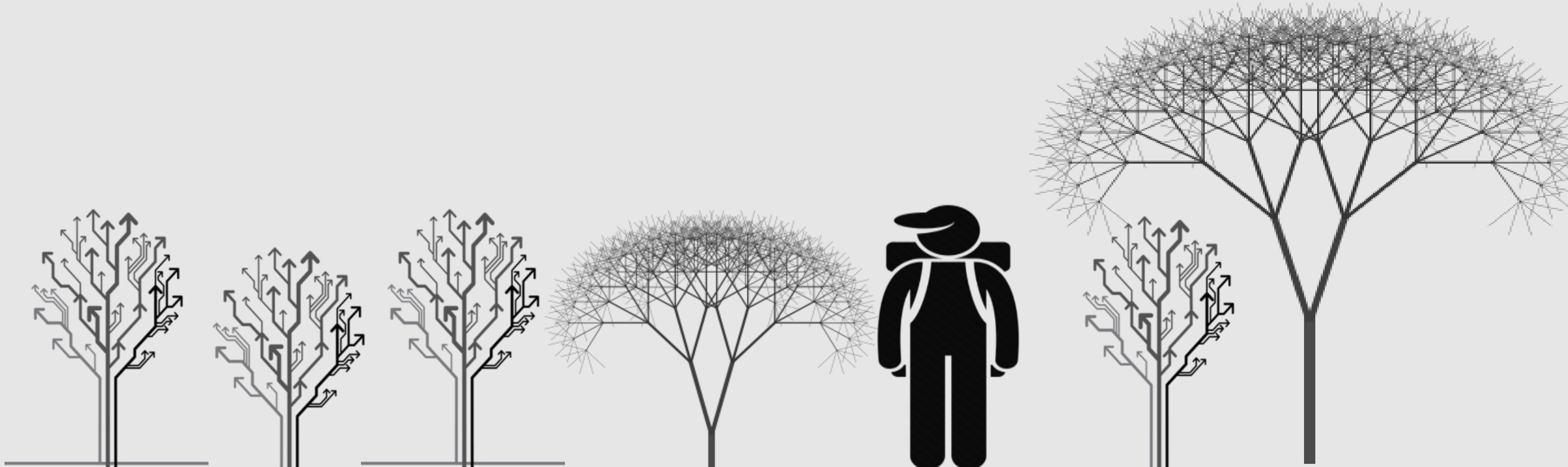
In Machine Learning:	In linear regression:
$Y_i = F(x_i) + \varepsilon_i$ Best fit $F(x_i)$ minimizes ε_i No assumptions about ε_i	$Y_i = F(x_i) + \varepsilon_i$ Best fit $F(x_i)$ minimizes ε_i Assume normal distribution of ε_i
Numerical approximation	Closed functional form



Why

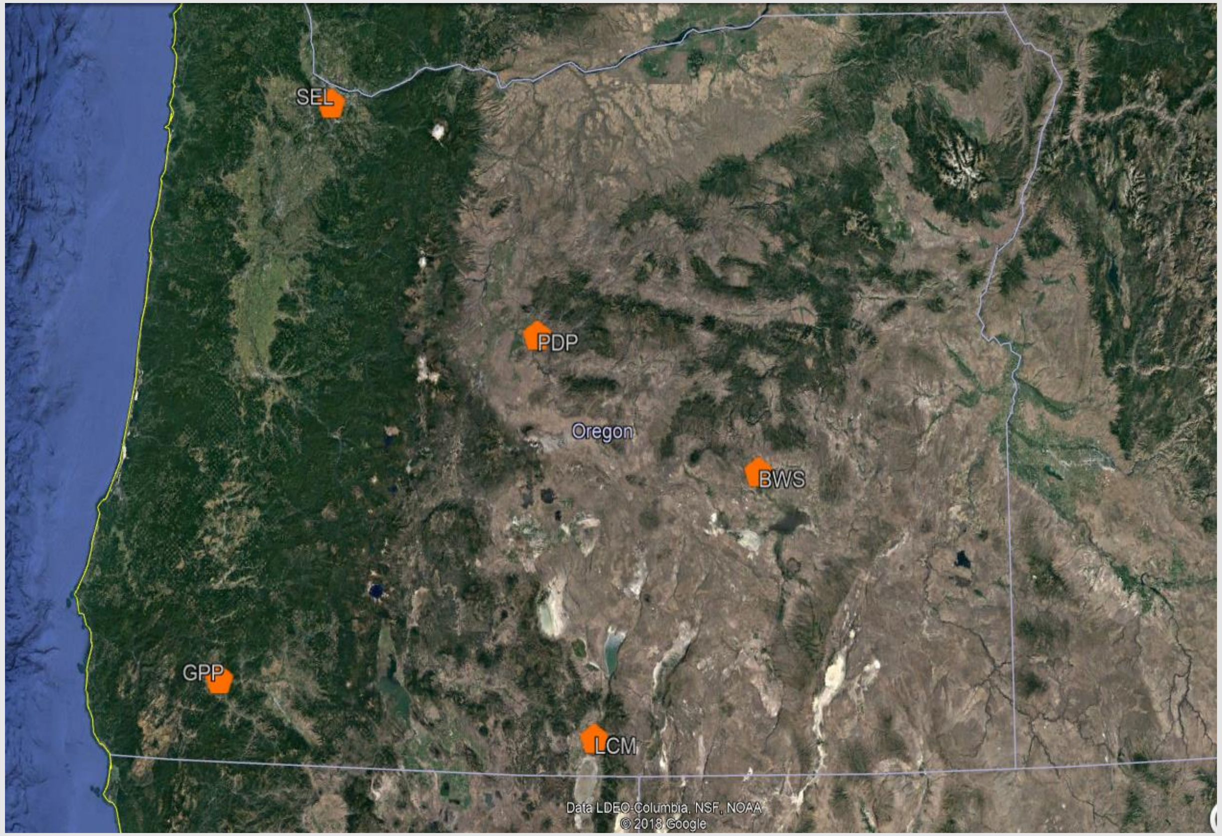
Machine Learning?

- No *a priori* distribution requirements
- Can handle correlated predictors
- Can potentially handle p > n
- Can handle multiple outputs



Our Exploration Plan

Predict the Nowcast



Five monitoring sites

Three ML algorithms

- Random forest
- Generalized boosted models
- Multi-layer perceptron

Five years of hourly PM2.5 data

- 2011 – 2014 training data
- 2015 validation data

Evaluation

- Three models
- Comparison of predictions to observations
  - Goodness of fit: R^2
  - Error: RMSE
- Comparison of predictions to Reff Nowcast

PREDICT THE NOWCAST

Predict NOWCAST: current hour PM2.5

M1

PM2.5 ~ pm25-1 + pm25-2 +pm25-3 + pm25-4 + pm25-5 + pm25-6

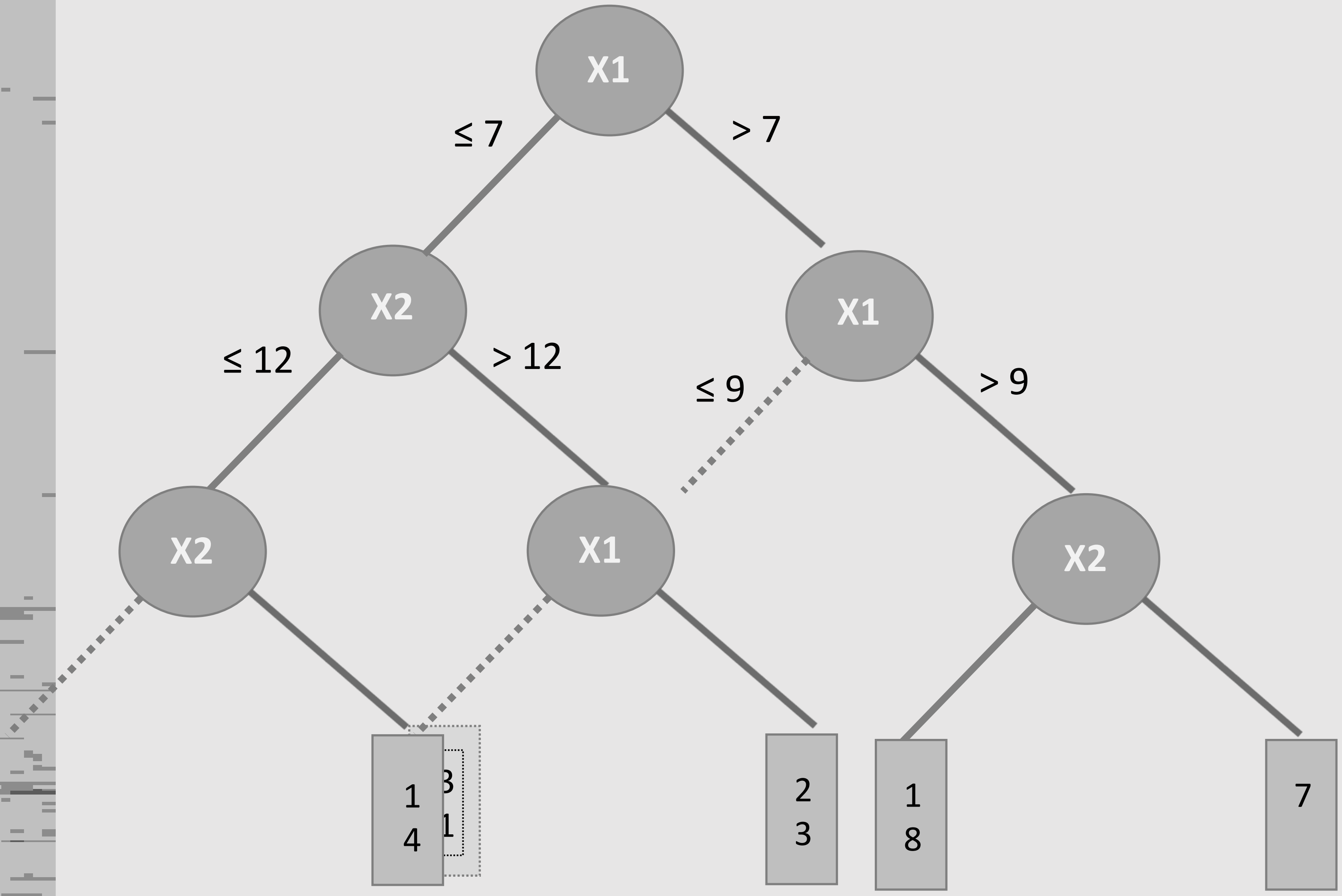
M2

PM2.5 ~ pm25-1 + pm25-2 +pm25-3 + pm25-4 + pm25-5 + pm25-6
+ hour + weekday + month

M3

PM2.5 ~ pm25-1 + pm25-2 +pm25-3 + pm25-4 + pm25-5 + pm25-6
+ hour + weekday + month
+ temperature + wind speed + wind direction

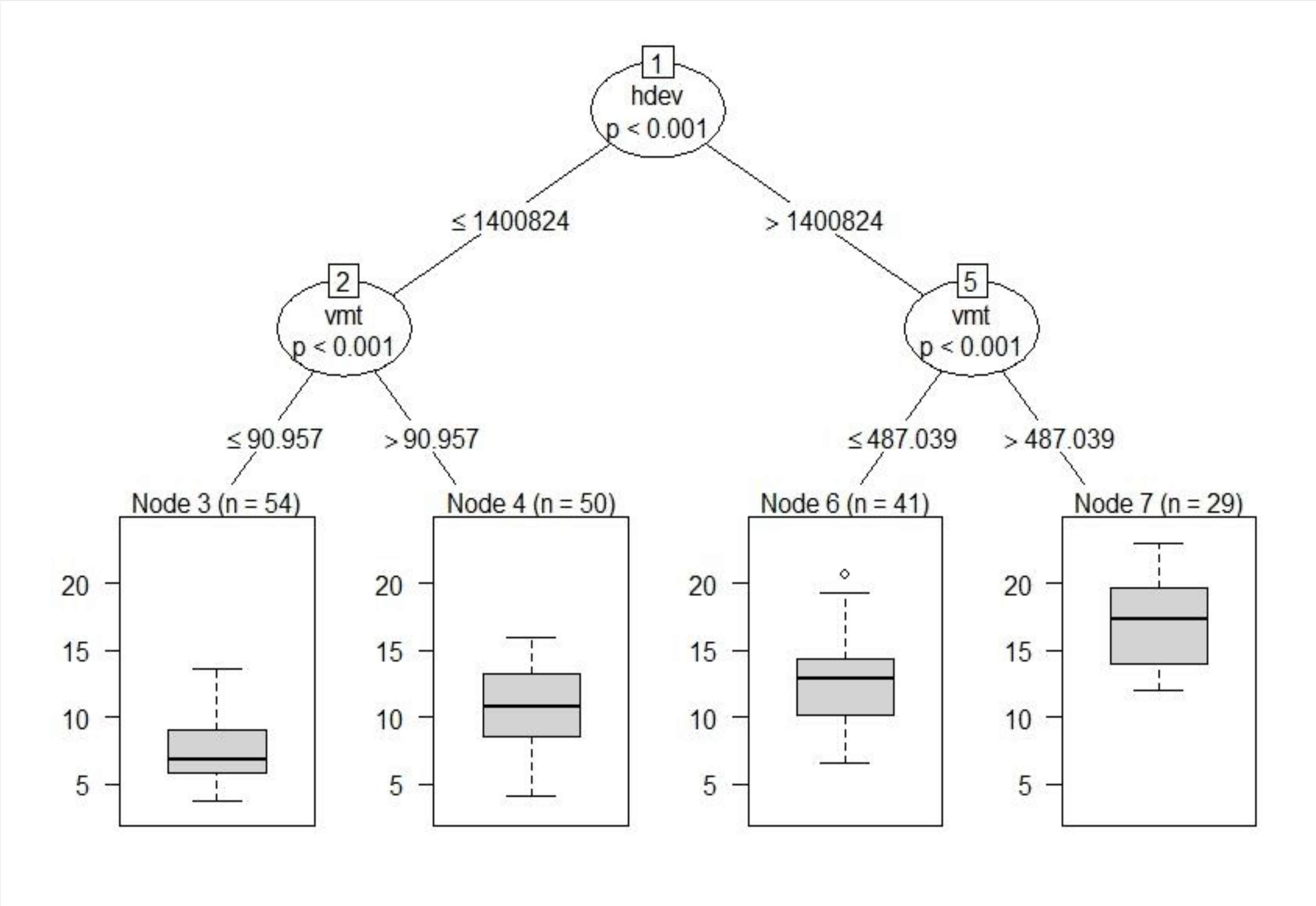
REGRESSION TREES



100 observations
Y – variable of interest
X1, X2 – predictors

$0 \le X1 \le 10$
$0 \le X2 \le 20$

Regression Trees



Predict:
- air pollution

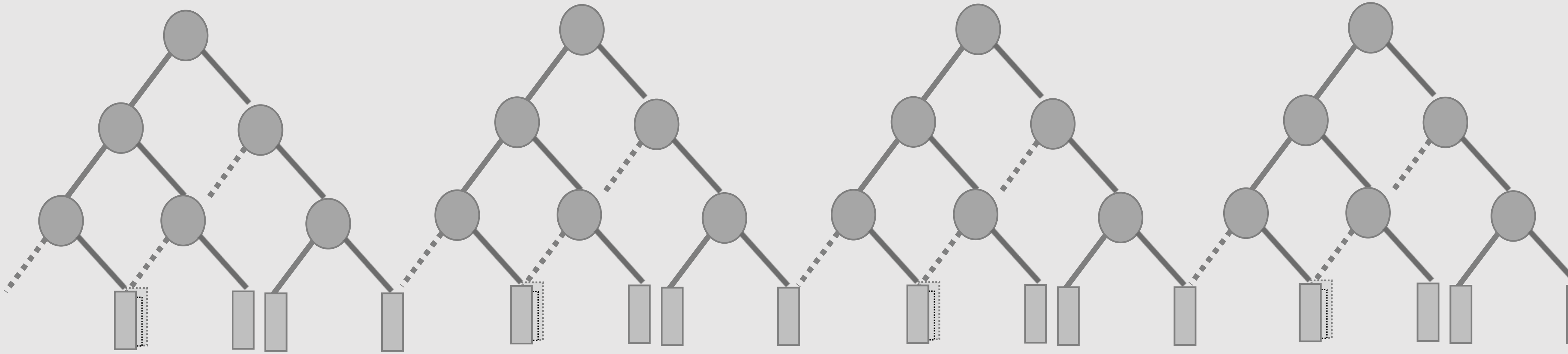
Predictors:
- hdev: high intensity development
- VMT: vehicle miles traveled

Number of observations:
- 174

An ensemble of *REGRESSION TREES*

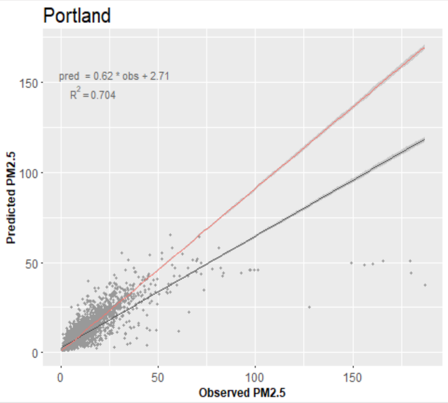
RANDOM FOREST

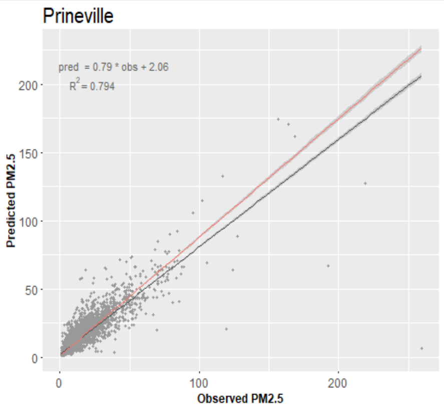
- Developed by Breiman (2001)
- Combine many “weak learners” into a “strong learner”
- Use bootstrap aggregation or **bagging**
- Each tree uses only a **random** subset of predictors

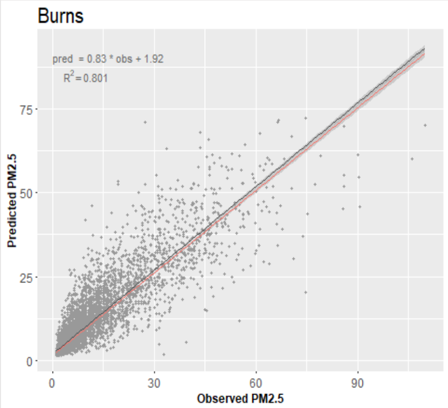


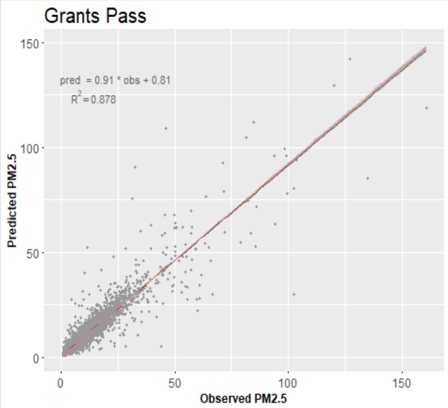
Random Forest Results

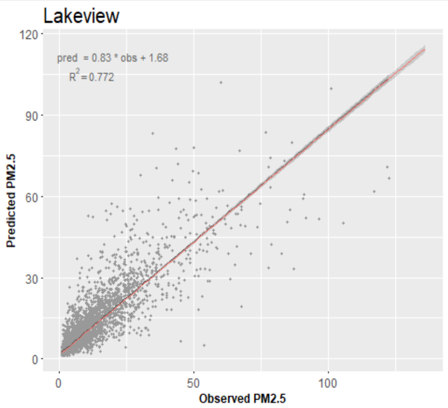
		REFF	RF1	RF2	RF3
Portland	R^2	0.83	0.73	0.73	0.70
	RMSE	3.8	4.8	4.8	5.0
Prineville	R^2	0.77	0.76	0.78	0.79
	RMSE	6.0	5.9	5.7	5.5
Burns	R^2	0.70	0.71	0.74	0.80
	RMSE	6.0	5.8	5.4	4.8
Grants Pass	R^2	0.86	0.88	0.88	0.86
	RMSE	3.3	3.1	3.0	3.1
Lakeview	R^2	0.71	0.71	0.74	0.77
	RMSE	5.3	5.4	5.0	4.7

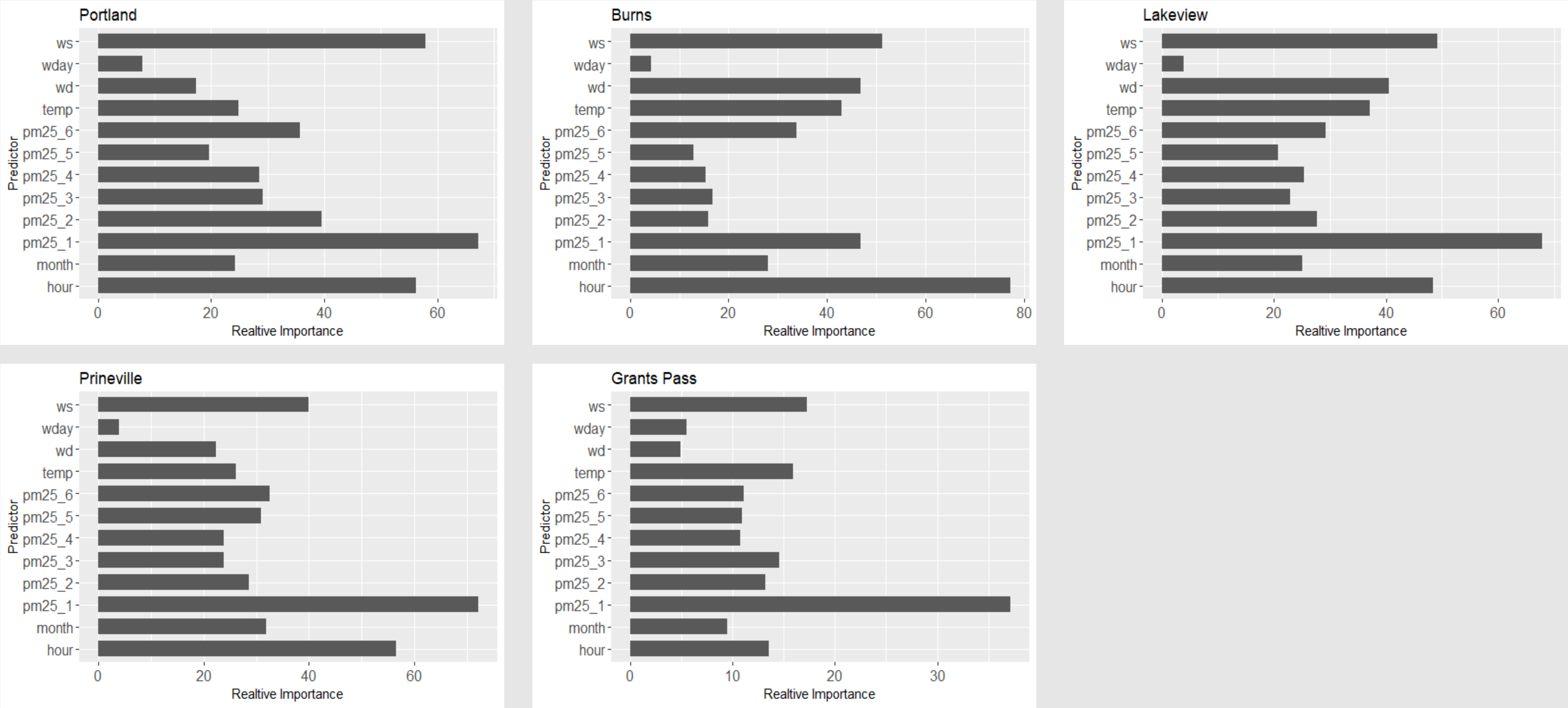


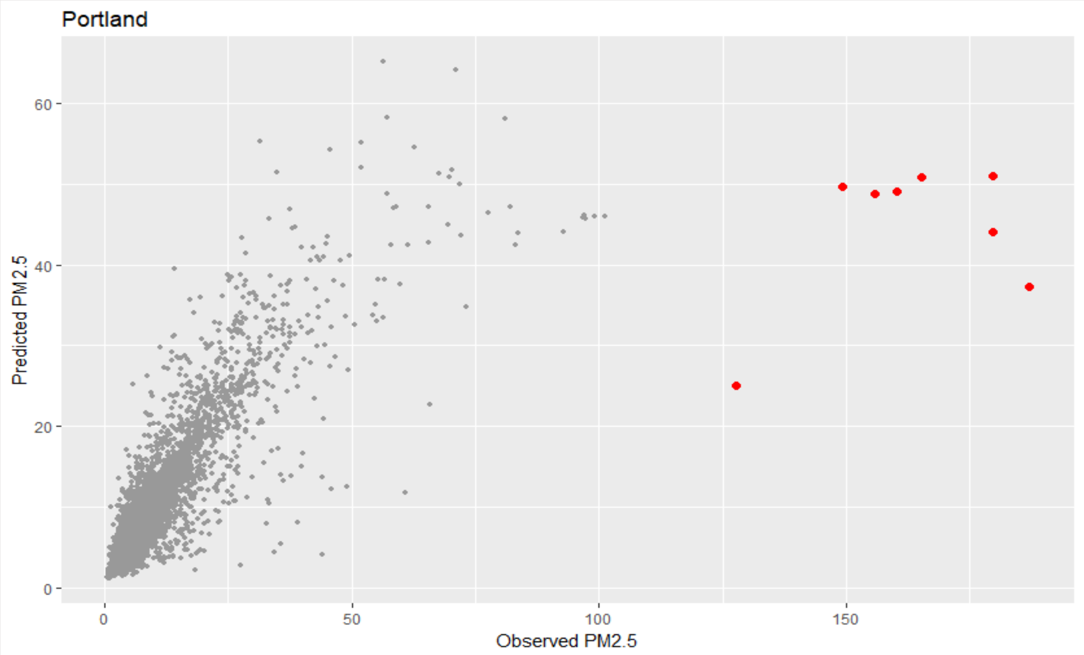




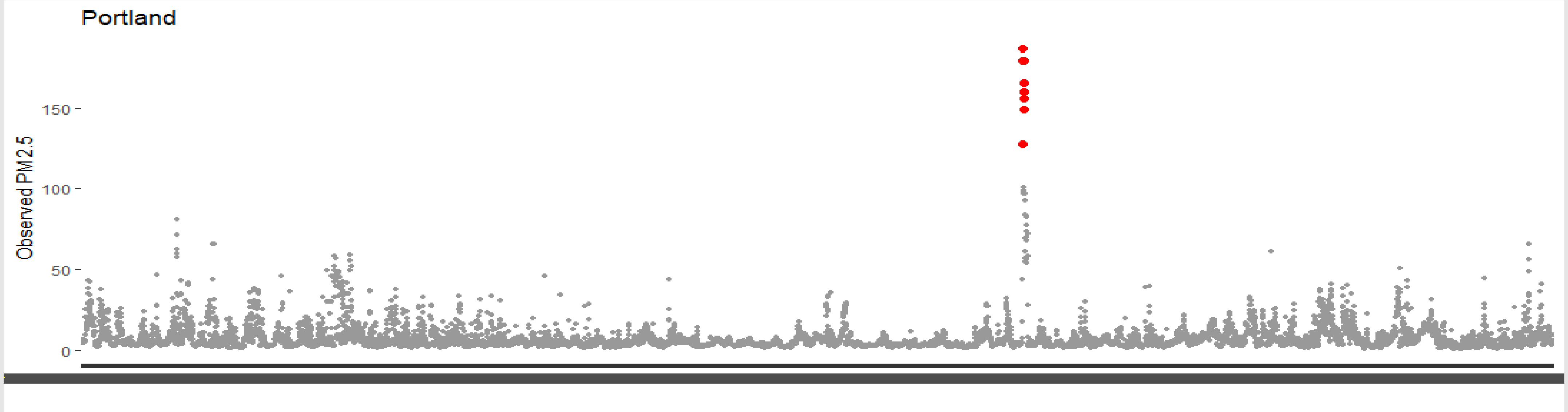





Portland
Burns
Lakeview
Prineville
Grants Pass
ws
wday
wd
temp
pm25_6
pm25_5
pm25_4
pm25_3
pm25_2
pm25_1
month
hour
Predictor
Realtive Importance
0
10
20
30
40
60
80



date	time	obs	reff	rf3
8/22/2015	11:00:00	127.6	42.2	25.0
8/22/2015	12:00:00	186.9	125.6	37.3
8/22/2015	13:00:00	179.7	185.8	44.1
8/22/2015	14:00:00	179.6	179.8	50.9
8/22/2015	15:00:00	165.3	179.6	50.8
8/22/2015	16:00:00	149.2	165.6	49.6
8/22/2015	17:00:00	155.9	149.6	48.7
8/22/2015	18:00:00	160.3	155.7	49.0



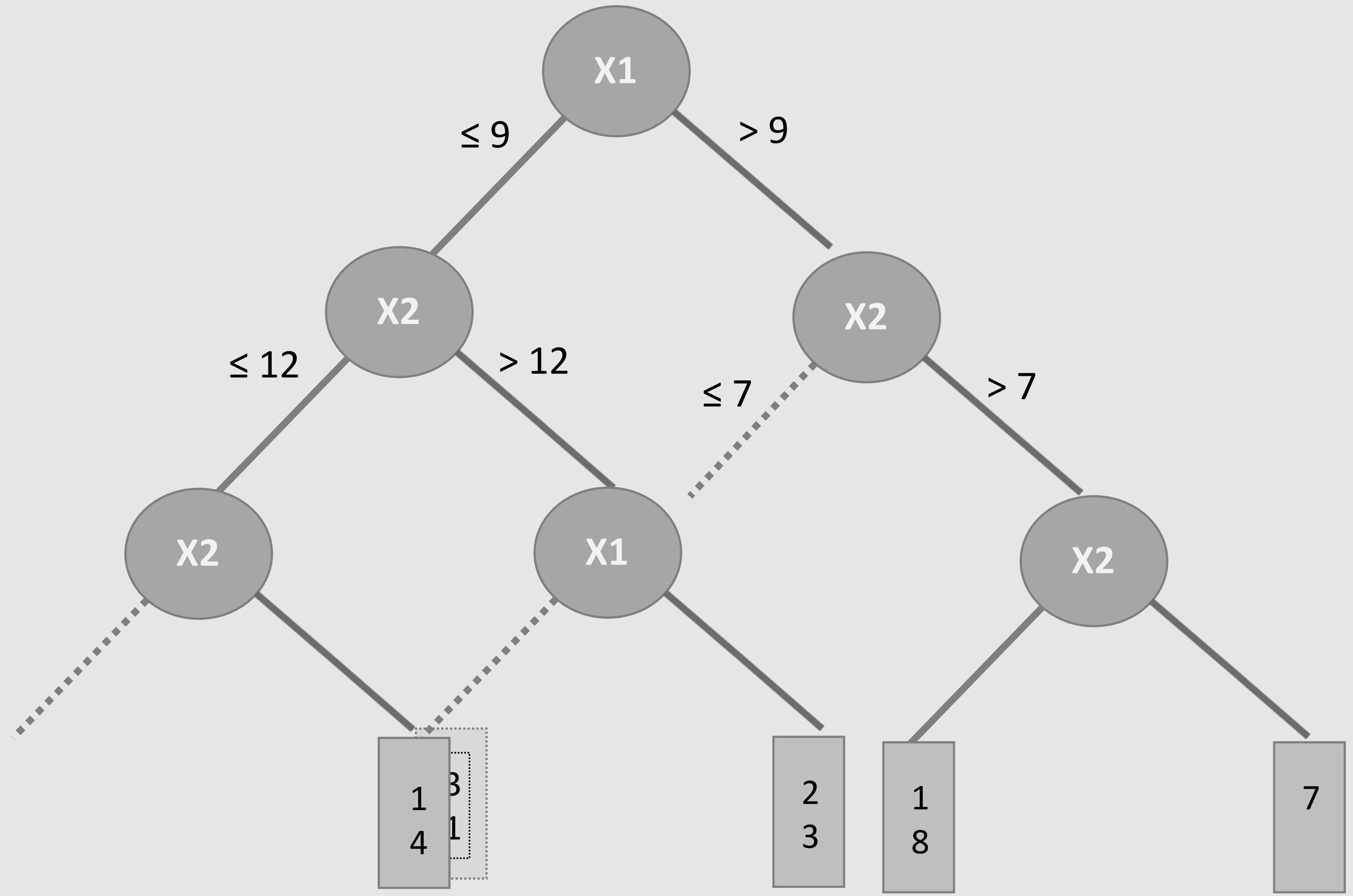
Air Quality Slips To 'Unhealthy' Levels Due To Wildfire Smoke
by **OPB staff** OPB Aug. 22, 2015 11:30 a.m. | Updated: Aug. 23, 2015 8:08 a.m. | Portland

" **...**Portland Fire & Rescue said it received numerous calls from residents reporting smoke in the area. Smoke is expected to increase throughout the day as winds travel approximately 26 miles per hour from east to west.
The smoke has blown from the Cougar Creek Fire near Mt. Adams, and the 12 other large fires burning east of the Washington Cascades..."

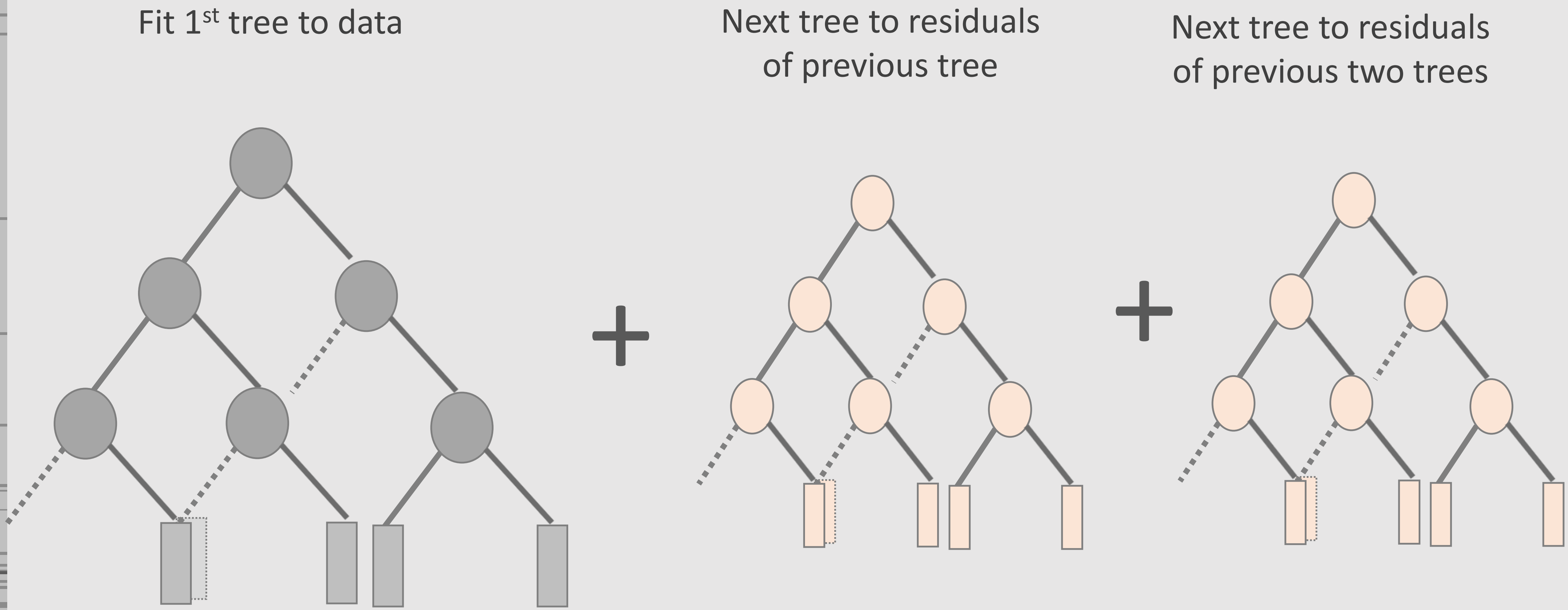
2011 to 2014 –
highest hourly PM2.5 was 94 μg/m^3
2015 –
highest hourly PM2.5 was 187 μg/m^3

GENERALIZED BOOSTED

REGRESSION MODELS

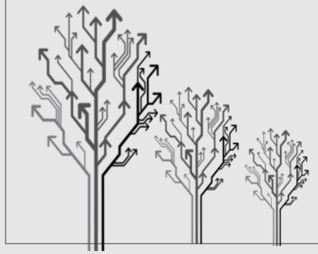


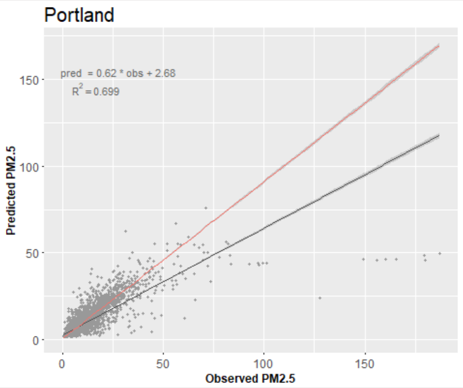
Generalized Boosted Regression Models

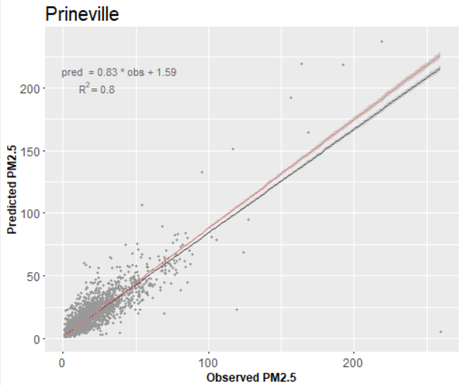


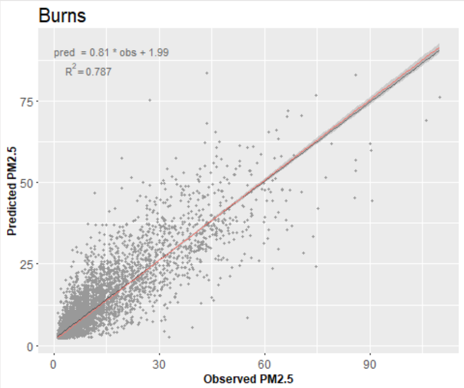
		REFF	GBM1	GBM2	GBM3
Portland	R^2	0.83	0.76	0.76	0.70
	RMSE	3.8	4.5	4.5	5.1
Prineville	R^2	0.77	0.77	0.78	0.80
	RMSE	6.0	5.8	5.7	5.4
Burns	R^2	0.70	0.72	0.74	0.79
	RMSE	6.0	5.6	5.4	4.9
Grants Pass	R^2	0.86	0.87	0.87	0.87
	RMSE	3.3	3.1	3.1	3.1
Lakeview	R^2	0.71	0.72	0.73	0.75
	RMSE	5.3	5.1	5.0	4.8

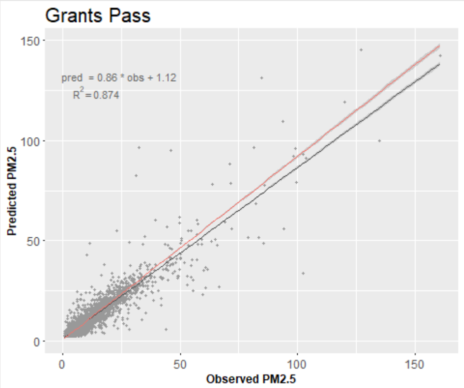
GBM RESULTS

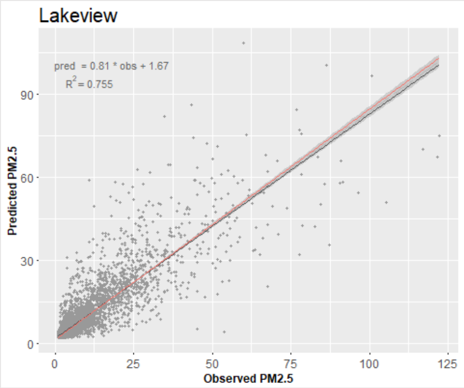






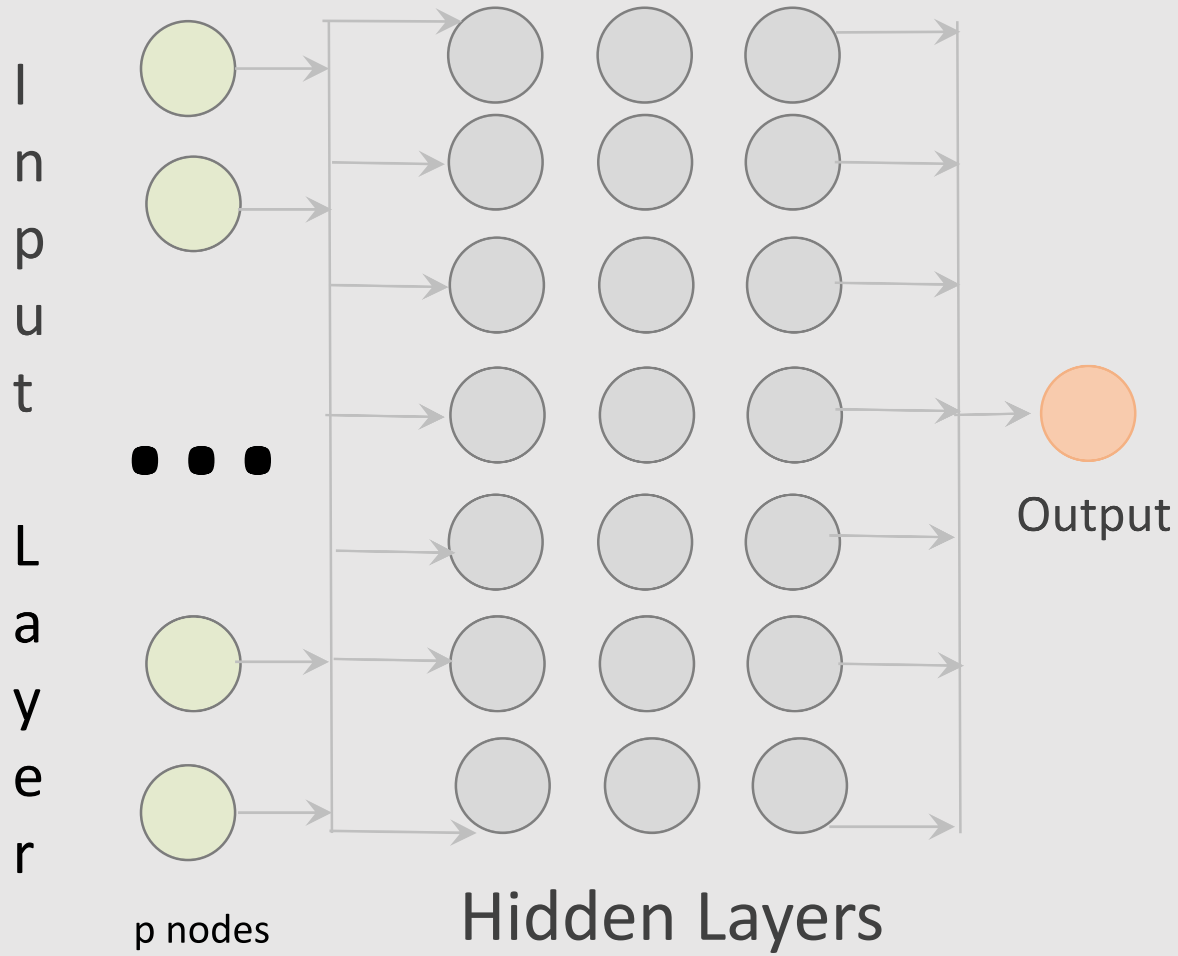


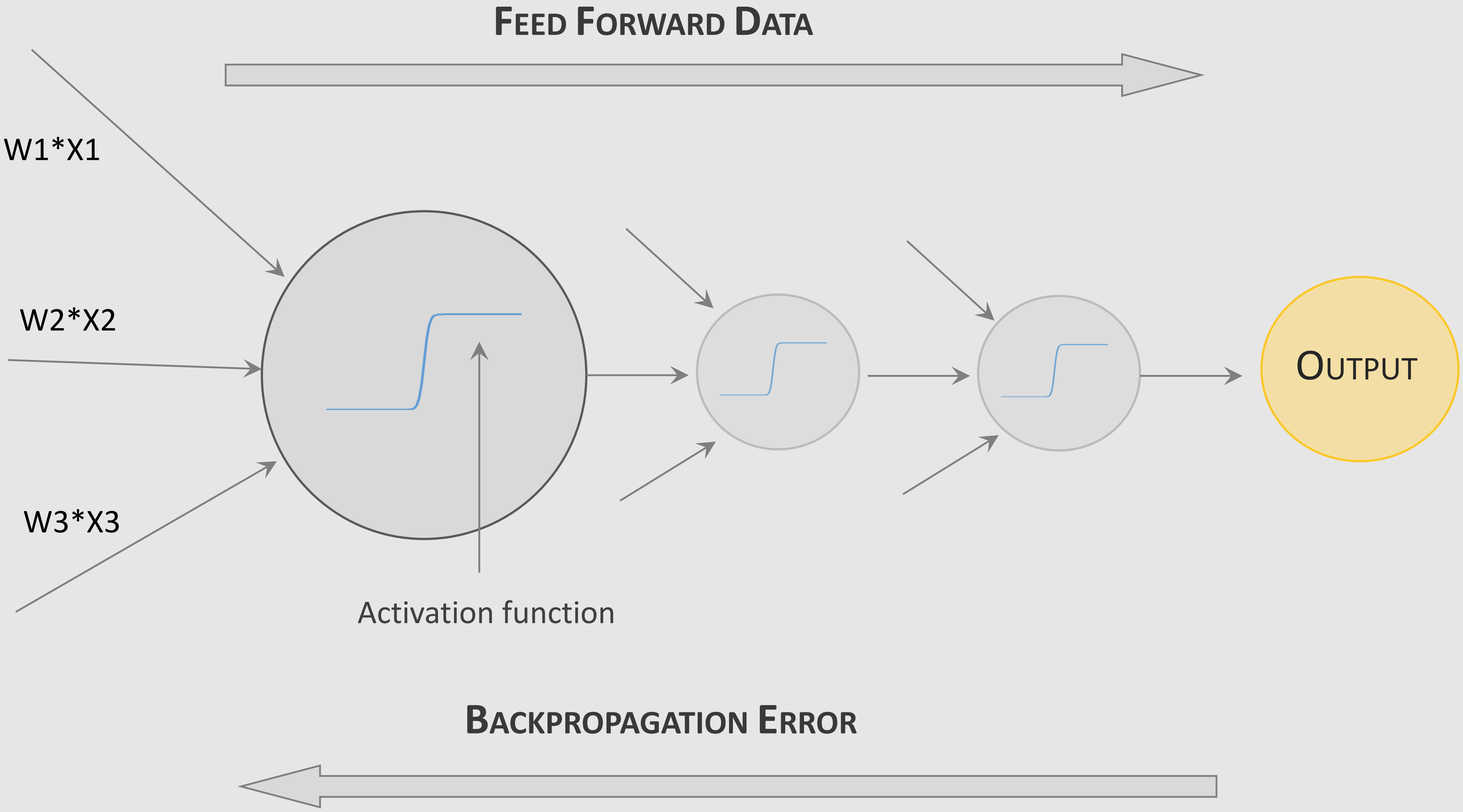




NEURAL NETWORK

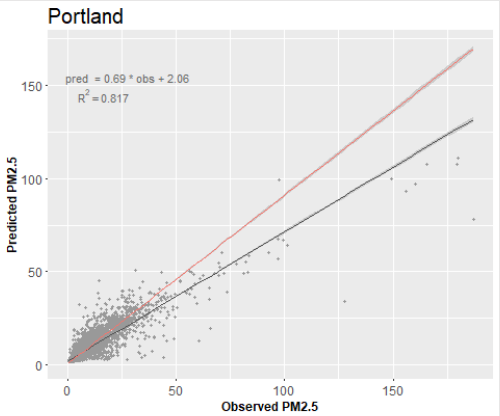
MULTI-LAYER PERCEPTRON

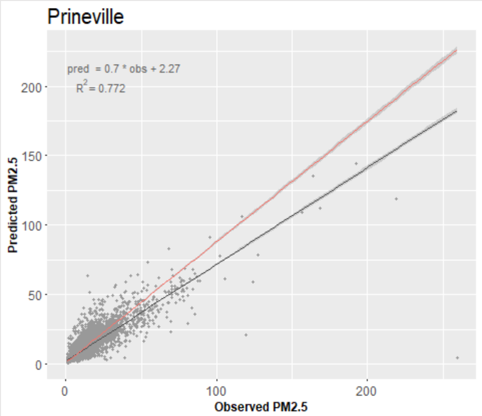


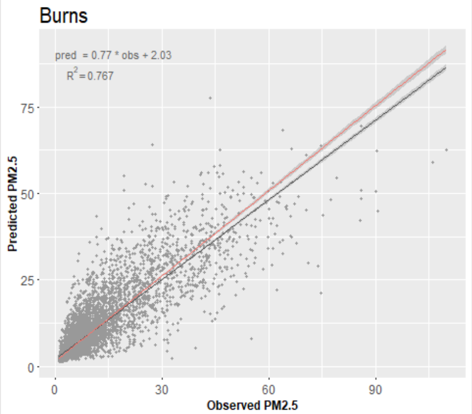

FEED FORWARD DATA
W1*X1
W2*X2
W3*X3
Activation function
OUTPUT
BACKPROPAGATION ERROR

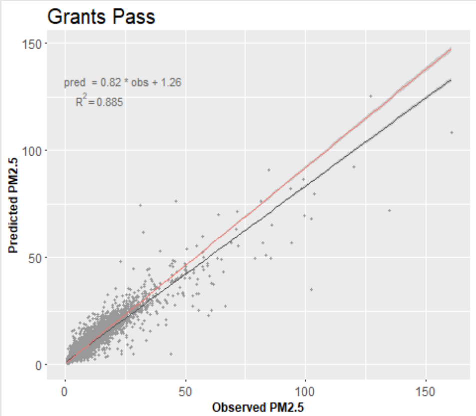
MLP Results

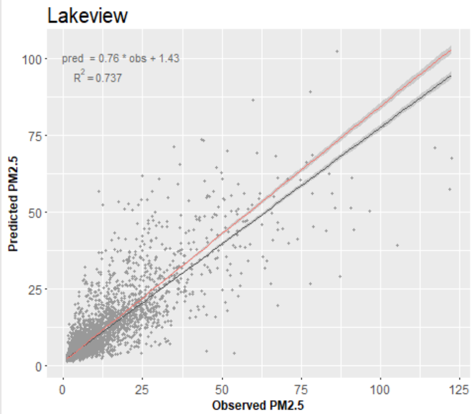
		REFF	MLP1	MLP2	MLP3
Portland	R^2	0.83	0.81	0.82	0.82
	RMSE	3.8	4.0	3.9	4.1
Prineville	R^2	0.77	0.77	0.76	0.77
	RMSE	6.0	5.9	5.9	5.9
Burns	R^2	0.70	0.71	0.73	0.77
	RMSE	6.0	5.7	5.6	5.1
Grants Pass	R^2	0.86	0.89	0.89	0.89
	RMSE	3.3	3.0	3.0	3.0
Lakeview	R^2	0.71	0.72	0.73	0.74
	RMSE	5.3	5.1	5.1	5.0











		REFF	RF3	GBM3	MLP3
Portland	R^2	0.83	0.70	0.70	0.82
	RMSE	3.8	5.0	5.1	4.1
Prineville	R^2	0.77	0.79	0.80	0.77
	RMSE	6.0	5.5	5.4	5.9
Burns	R^2	0.70	0.80	0.79	0.77
	RMSE	6.0	4.8	4.9	5.1
Grants Pass	R^2	0.86	0.86	0.87	0.89
	RMSE	3.3	3.1	3.1	3.0
Lakeview	R^2	0.71	0.77	0.75	0.74
	RMSE	5.3	4.7	4.8	5.0

Random Forest Results

GBM Results

MLP Results

Operational Details

Training N: ~ 35,000
Validation N: ~ 8,700
Predictors: 6 – 13
R: randomForest, gbm, keras

	RF	GBM	MLP
Time	~10 min	~6min	< 1min
Parallelization	No	Yes	Probably
Hyper-parameters	mtry, ntrees	n.trees, interaction.depth, n.minobsinnode, shrinkage, bag.fraction, train.fraction, cv.folds	Hidden layers, nodes in layer, activation functions, loss function, learning rate
Tuning & diagnostics	4	1	1
Insight	3	1	1

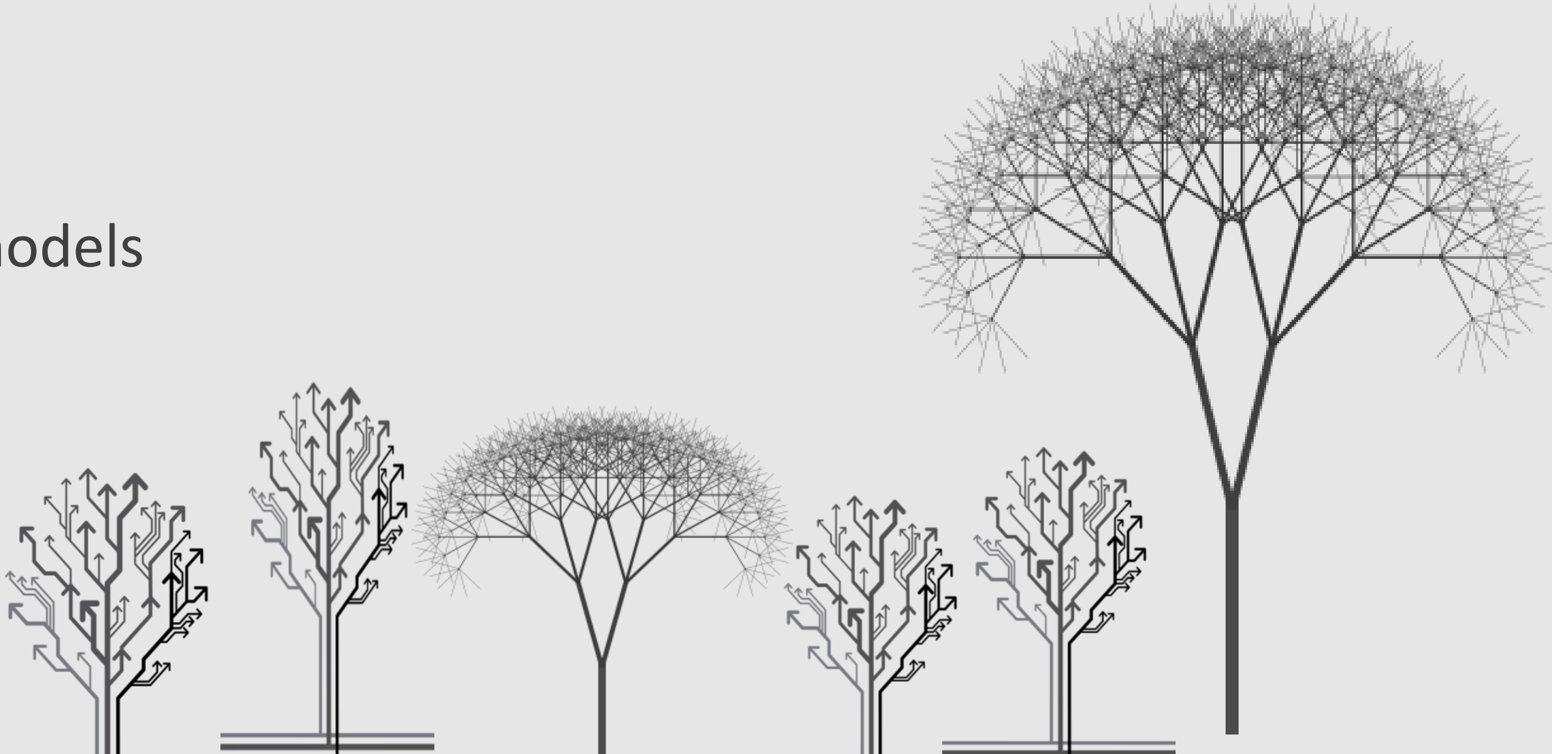
Our Learning About

Machine Learning

- Performed reasonably well
- Tuning and diagnostics techniques and tools still in infancy
- Tools to peer into the ML “blackbox” lacking

And…

- Multiple outputs
- Dynamically updated models
- Wave of the future



Thank You for Joining Us in this Exploration!