Supporting Information for "Toward data-driven generation and evaluation of model structure for integrated representations of human behavior in water resources systems"

Liam Ekblad[1], Jonathan D. Herman[1]

[1]Department of Civil and Environmental Engineering, University of California, Davis, CA, USA

Contents of this file

1. Table S1 - Feature data used in study

2. Figure S1 - Correlation heatmap for input data

3. Figure S2 - Model results accuracy vs. alternative metrics

4. Figure S3 - Model results complexity vs. alternative metrics

Land Feature Data		Water Feature Data	Economic Feature Data	
Tree Crops	Non-Tree Crops	Non-Ponded Crop Deliveries	Alfalfa	Almond
		Non-Ponded Crop Pumping	Apricot	Beeswax
		Rice Crop Deliveries	Cotton	Grape
		Rice Crop Pumping	Honey	Milk
		Urban Deliveries	Nectarine	Pistachio
		Urban Pumping	Plum	Walnut
		Refuge Deliveries	Wheat	
		Refuge Pumping		
		Total Pumping		

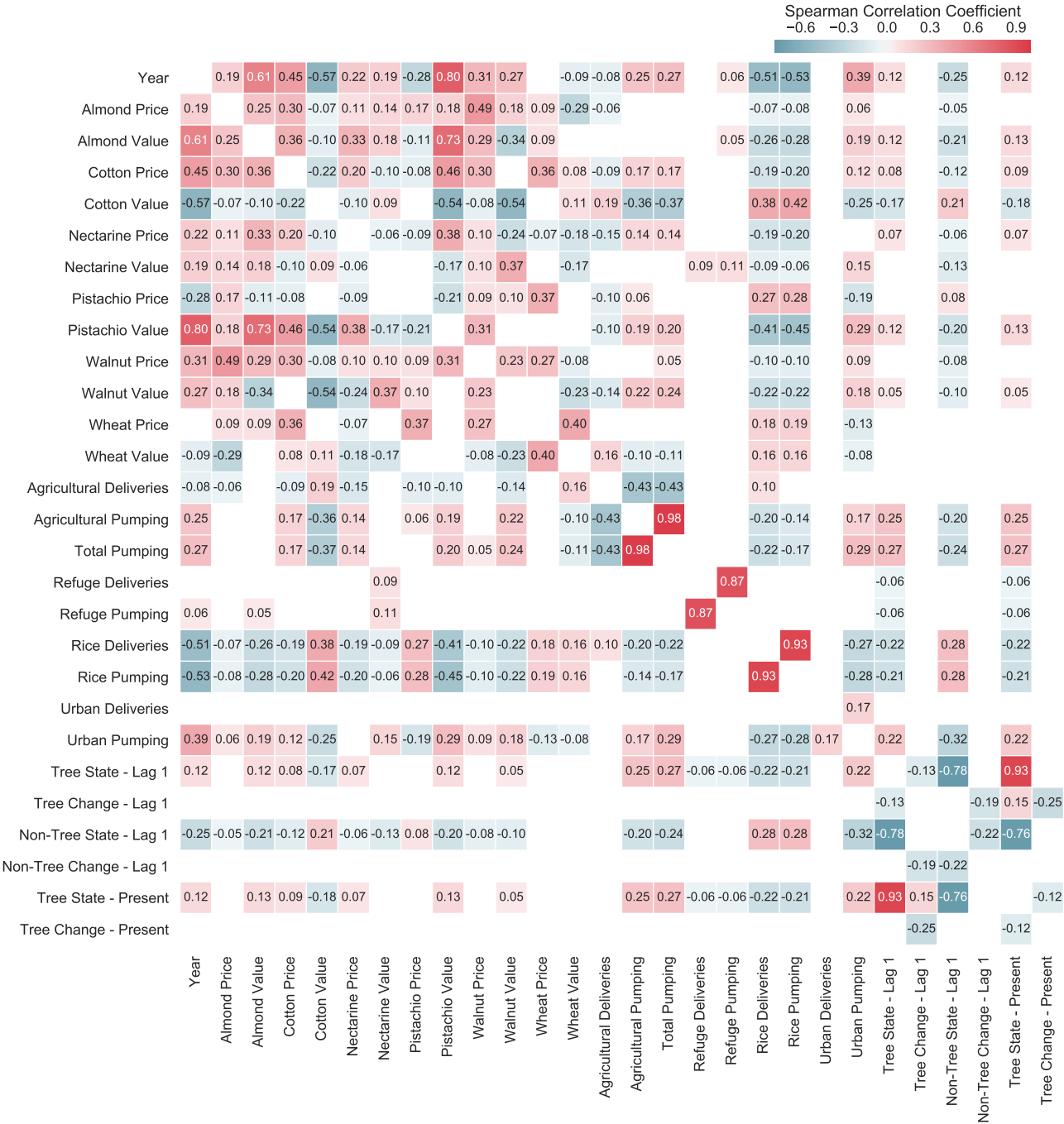**Table S1.** Feature data used to generate models during the experiment.

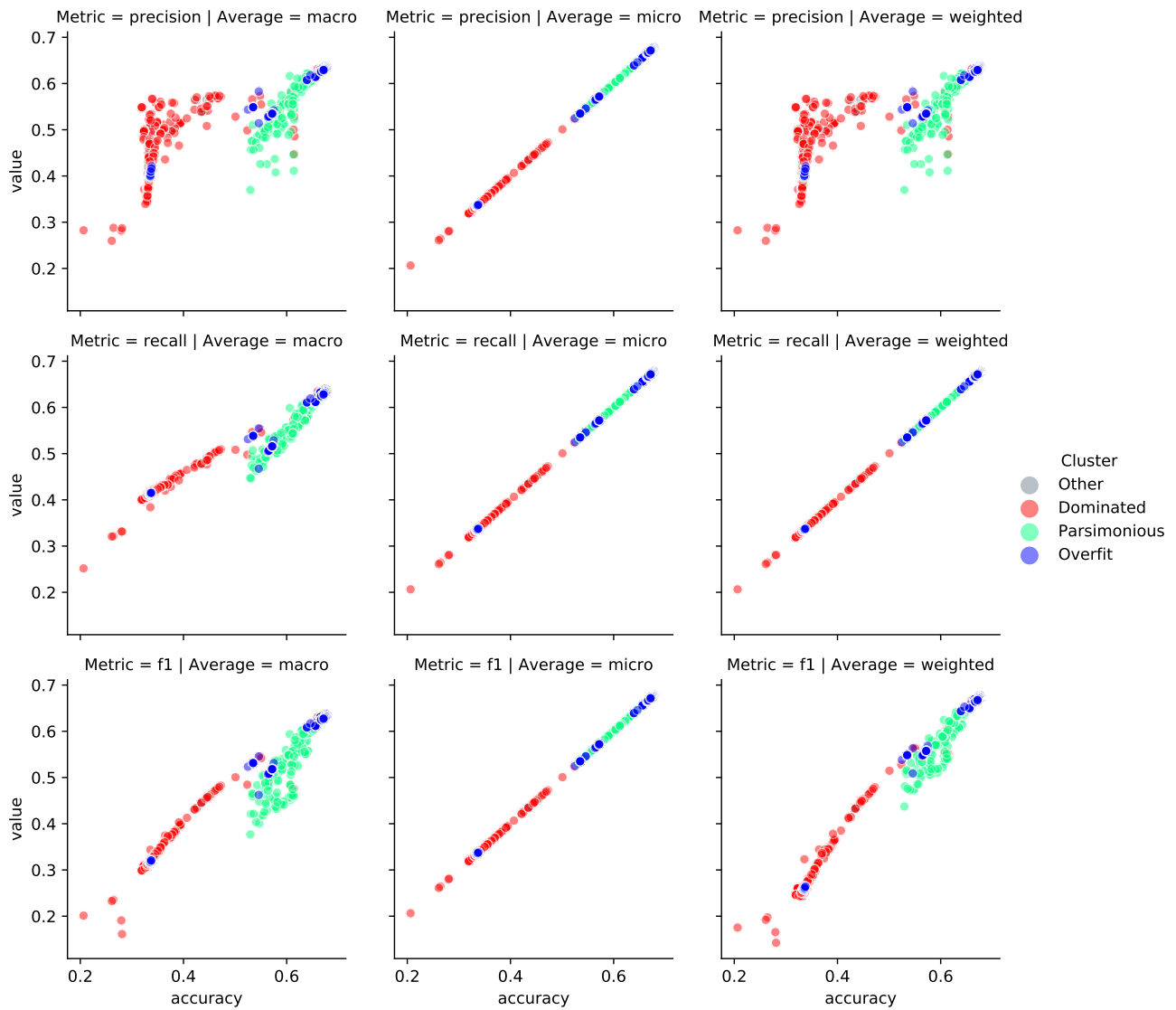**Figure S1.** Nonlinear correlations for a subset of the features represented in the table above.

Figure S2.   Visualization of classification model performance metrics in relation to the simple
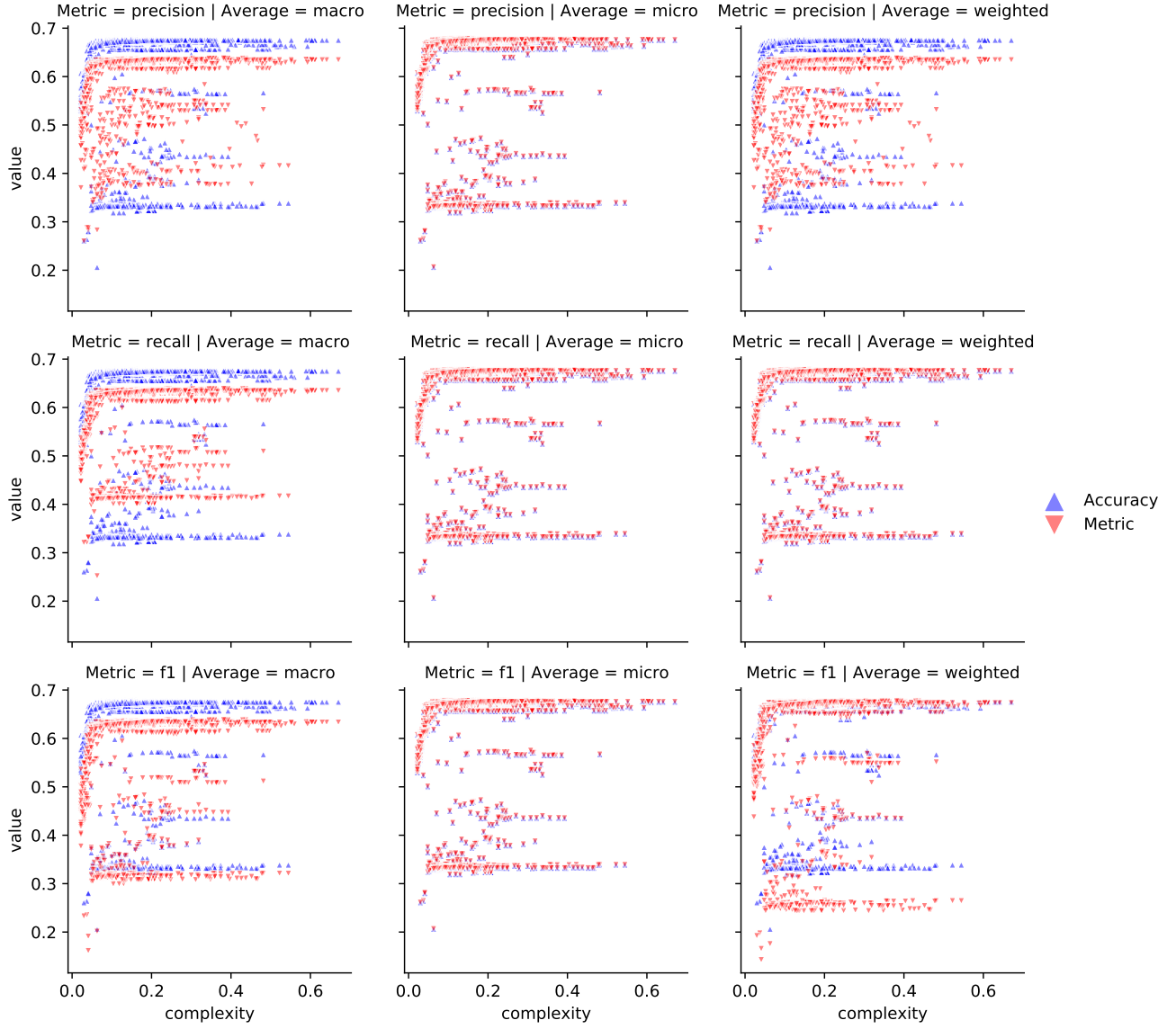
accuracy metric used in the paper.

Figure S3. Visualization of the effect of classification model performance metric selection on the resultant performance-complexity tradeoff.