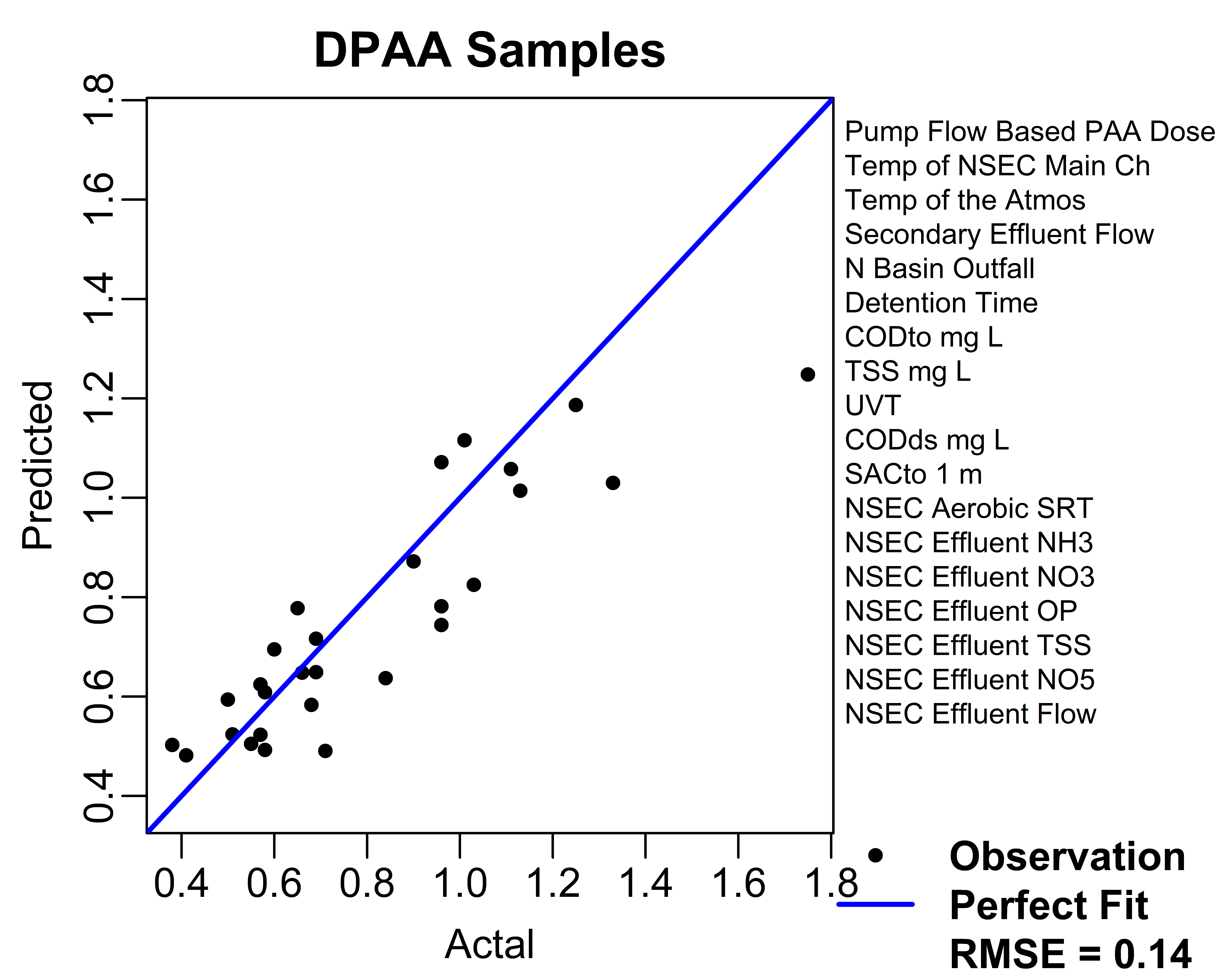
**Neural network model to predict PAA demand**

A mixture of online and grab data was used to predict the change in PAA throughout the North disinfection basin. To predict the total demand between initial dose and final concentration (at 1/2 basin), the following datasets were merged:

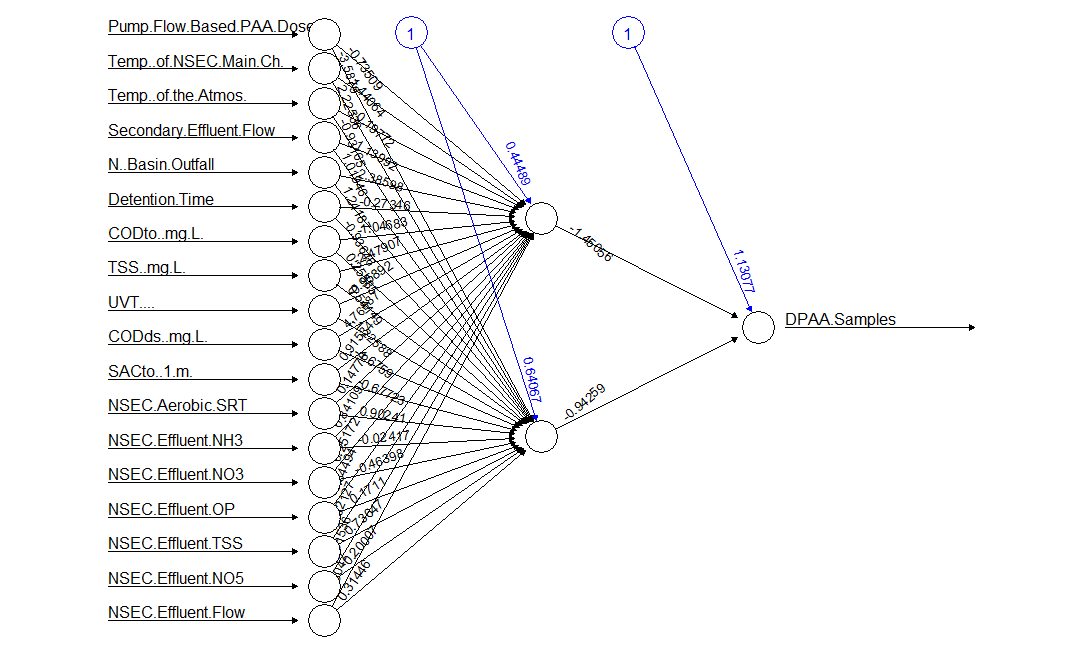
* 111 observations of PAA grab samples collected regularly between Oct 2-15, 2018
* Online data from a CarboVis analyzer upstream of PAA dosing
* Online data from the north secondary effluent

Data were used to train and test neural network (NN) models. A randomly selected 80% of the dataset was used for training the NN model. The remaining 20% were used to validate the model (Figure 1). NN structure (Figure 2) was selected by minimizing the root-mean-squared-error (RMSE, standard deviation of prediction errors).

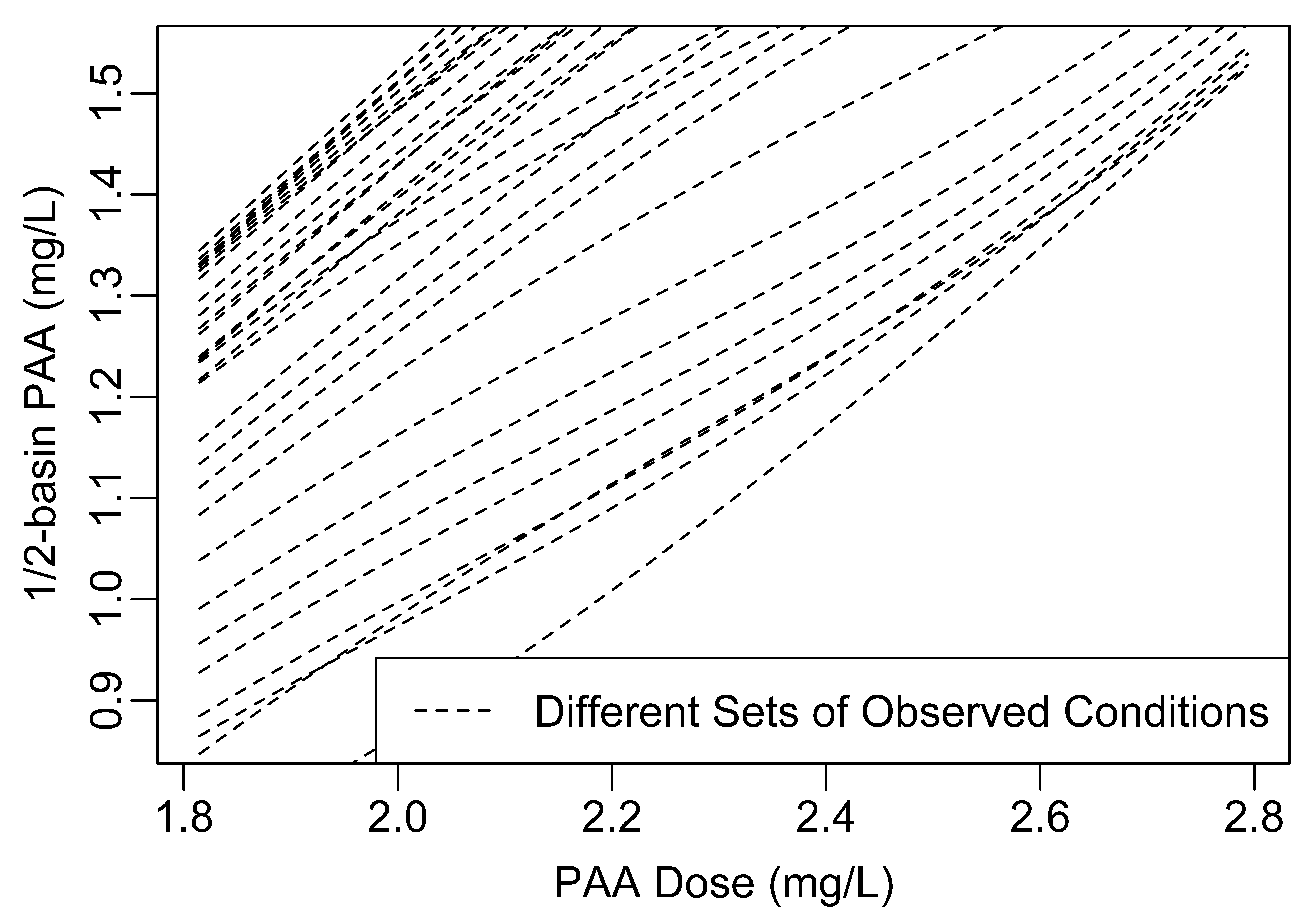
Ideally, the PAA pump flowrate could be adjusted to minimize the concentration of PAA at the end of the disinfection basin given a minimum CT. However, there is not a constant linear relationship between initial PAA dose and the ½ basin concentration (Figure 3).



**Figure 1.** Neural network model prediction of the change in PAA concentration (“DPAA” or Delta PAA) between the 1-minute PAA grab sample and the ½ basin PAA grab sample. Black dots are test data that include DPAA and the process variables listed to the right of the plot for each observation (i.e., point). RMSE is the root-mean-squared-error between the predicted and actual DPAA values and was used to identify the optimum NN model. The blue line is representative of a perfect model fit (i.e., predicted = actual).



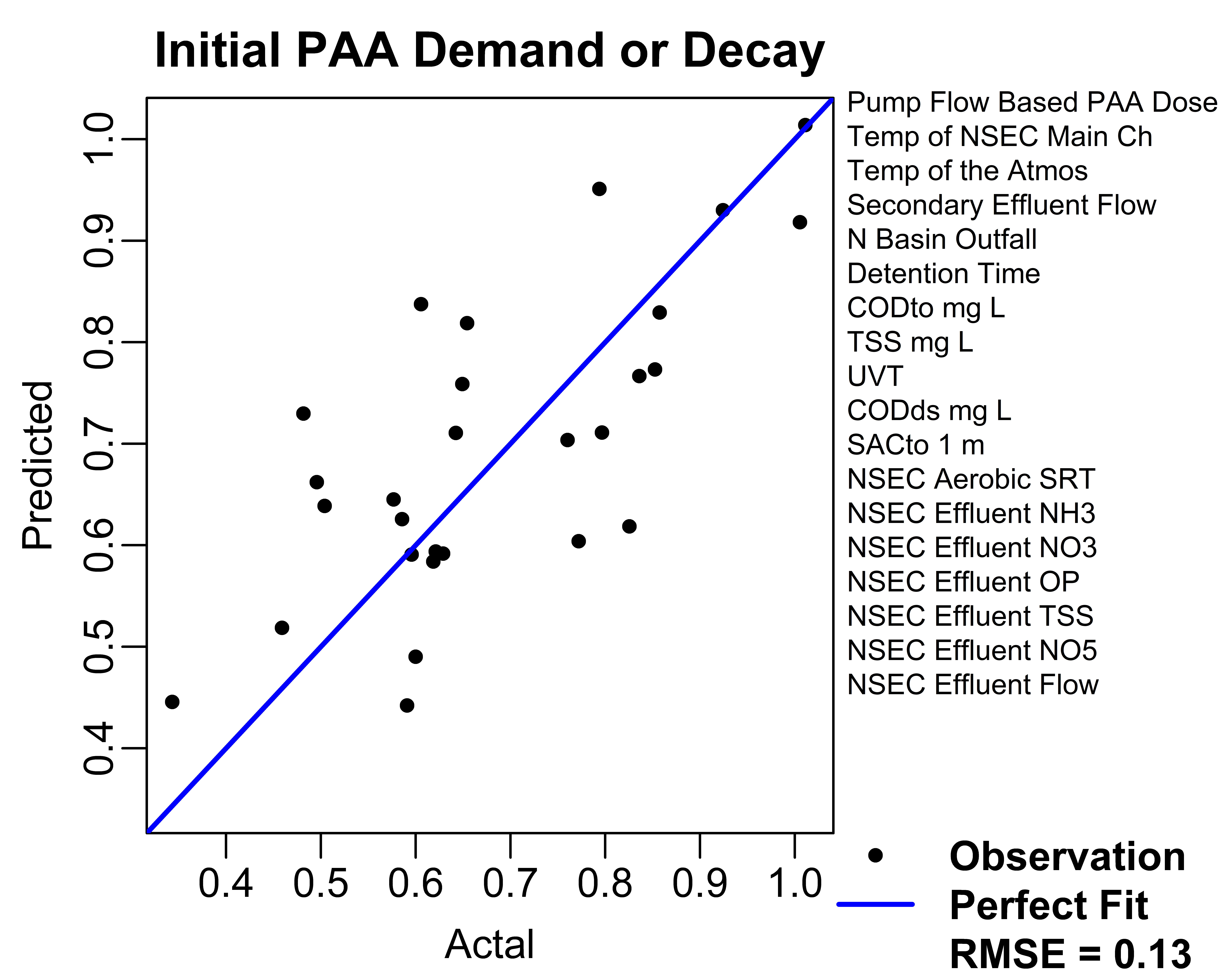
**Figure 2.** Example of the optimum NN structure used to calculate change in PAA concentration (“DPAA”). NN model was optimized by changing number of hidden nodes to minimize the RMSE; 2 are used in this example. At each node (circle), an activation function (sigmoid) transforms the weighted values (weights in black text) of each input plus a bias term (blue text). The weights and biases are determined by resilient back-propagation with weight backtracking.



**Figure 3.** For all test observations (datapoints in Figures 1): PAA dose was varied (x-axis) and ½ basin PAA concentration was predicted (y-axis) using the NN model from Figure 2. Plot demonstrates the impact of other water quality parameters considered in the NN model beyond initial PAA dosing rate.

**Neural network models – Grab vs Online PAA**

To compare how initial PAA demand is predicted, grab data from October 2018 (Figure 4) and online data from December 2018 (Figure 5) was used to build predictive NN models. The range of online data is limited, but functions well for prediction of immediate PAA demand.



**Figure 4.** Neural network model prediction of the change in PAA concentration between the flow-based PAA dose and the 1-minute PAA grab sample. Black dots are test data that include initial PAA demand and the process variables listed to the right of the plot for each observation (i.e., point). RMSE is the root-mean-squared-error between the predicted and actual PAA demand and was used to identify the optimum NN model. The blue line is representative of a perfect model fit (i.e., predicted = actual).