Rinse Over Run Report

RnD Team

Kunnal Kainth and Francisco Mendez

Overview of our algorithm

- We **split** data in 6 different subsets according to the recorded phases of the process in the test set and the corresponding recipe and train one model (random forest) for each subset.
- To predict the final rinse total turbidity liter we identified that given the measurements of the last phase recorded the other phases measurements are independent from the predicted target. (E.g. Process 20010 in the test set had measurements from all the phases, our algorithm only used the measurements from the acid phase)
- The caustic phase if the exception to the previous statement. If the caustic phase was the last phase recorded, the pre rinse phase measurement were considered for the model. This leads us to assume that given the caustic phase, the final rinse phase is not independent from the pre rinse phase. This is important for the process with recipe 11001.

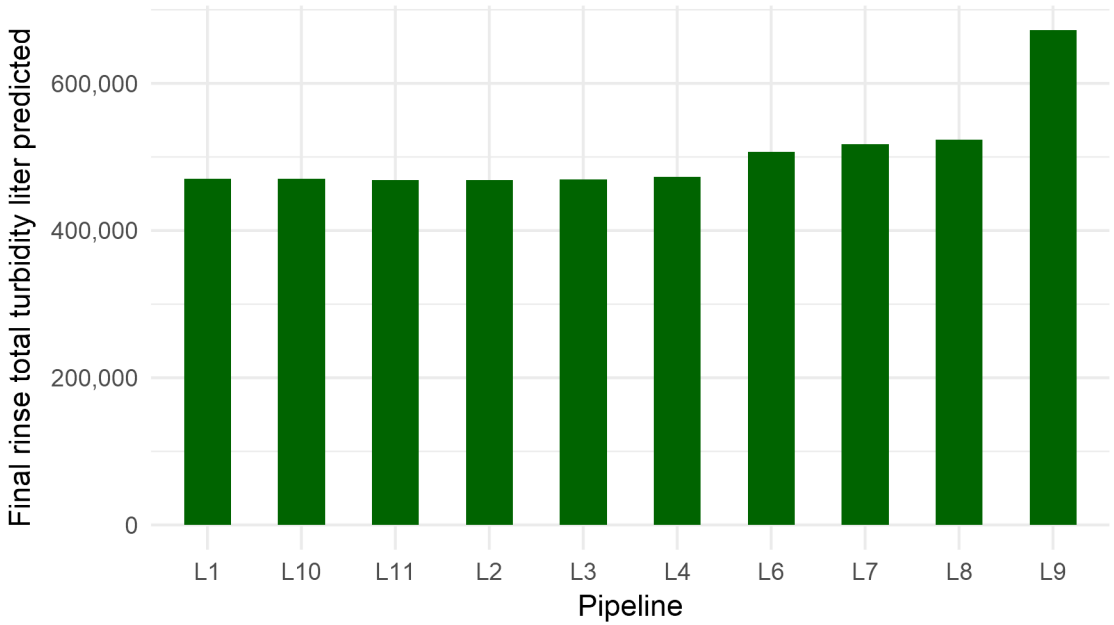
Sensitivity

- The impact of the variables presented on the following slides are under de assumption that the acid phase was observed.
- The following charts shows the prediction of a typical process (a process which values are the median of every variable) modifying only one variable at a time to highlight the impact of certain variables in the predictions.
- The variables shown below are the ones the model identified as the measures which impacted the most the prediction of the final rinse turbidity liter

Sensitivity

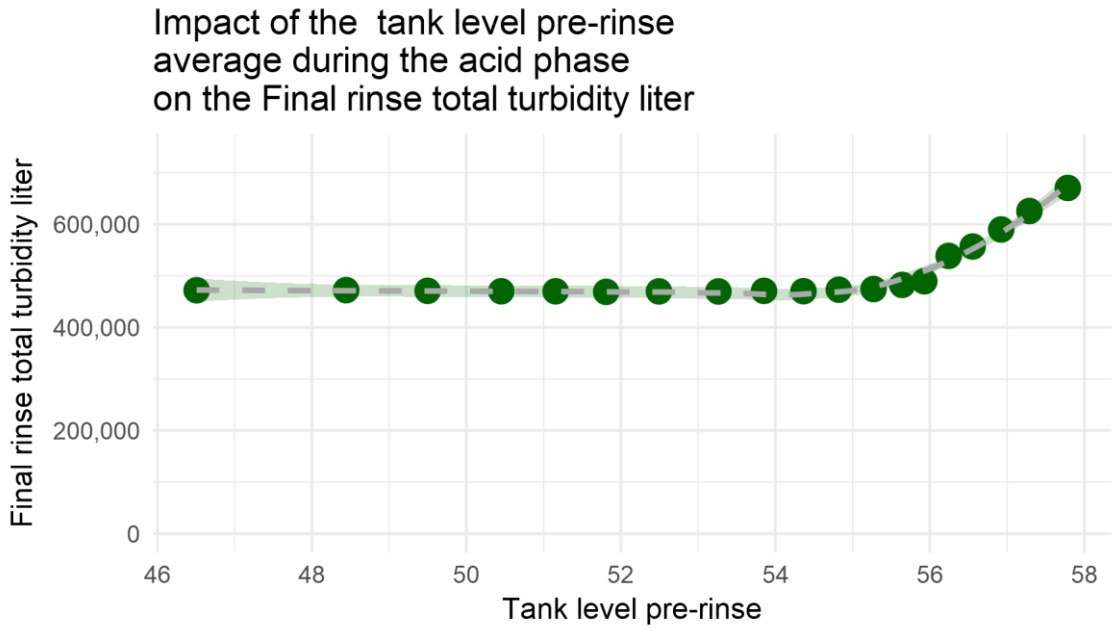
Pipelines

- Pipeline L9 increases by more than 40% the final rinse total turbidity liter of a similar process run in a different pipeline.



Tank level pre-rinse average

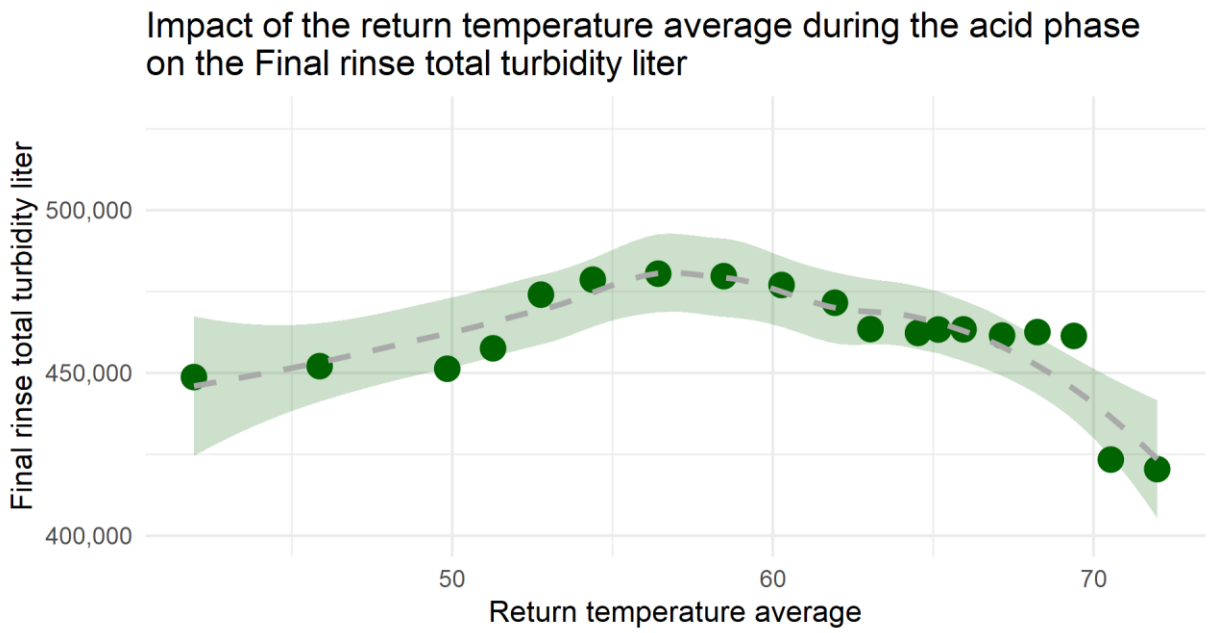
- The average of the tank level pre rinse average has a negative impact on the turbidity liter if it is bigger than 56. (More than 20% of the processes recorded a value higher than 56.)



Sensitivity

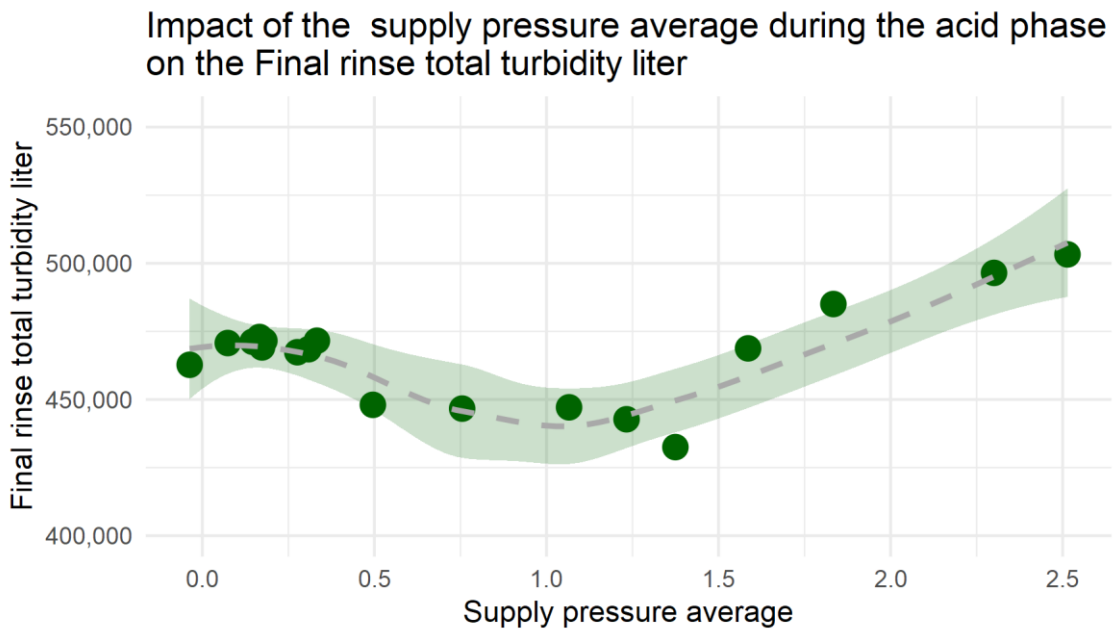
Return temperature average

- Higher return temperatures tend to decrease the final turbidity liter



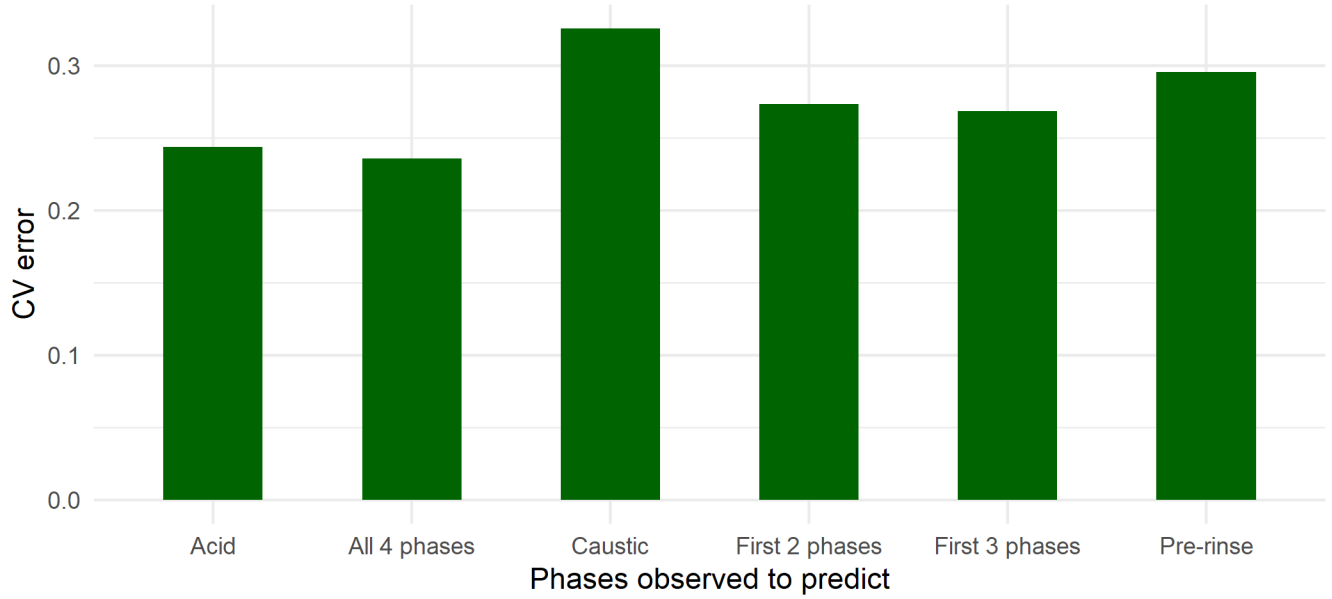
Supply pressure average

- High supply pressure during the acid phase tends to increase the final turbidity liter



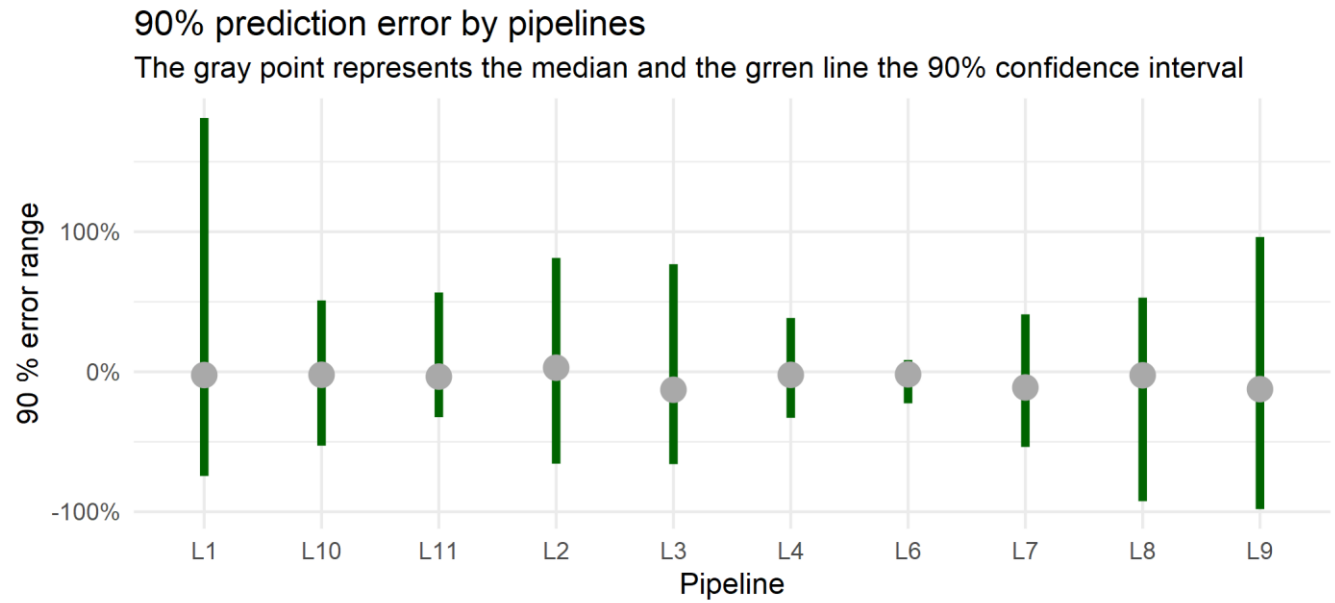
Confidence level of the predictions

- Our algorithm consisted in 6 different sub models, the prediction of each process was calculated using one of the 6 models depending on the observed phases. The CV errors of the prediction depending on the phases observed are shown below



Confidence level
Pipelines

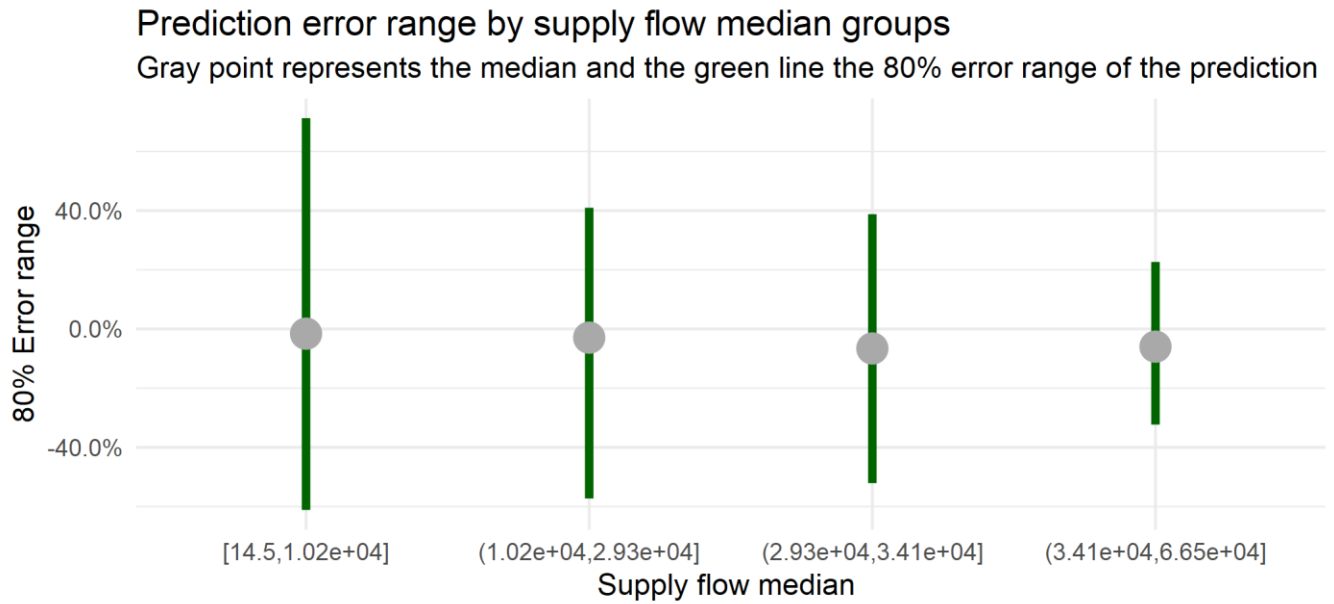
- Pipelines shown the major change in terms of predictive confidence, L1 tends to over predict while L9 tends to under predict. The error range of pipeline L6 and L4 are considerably smaller to the rest



Confidence level

Supply flow

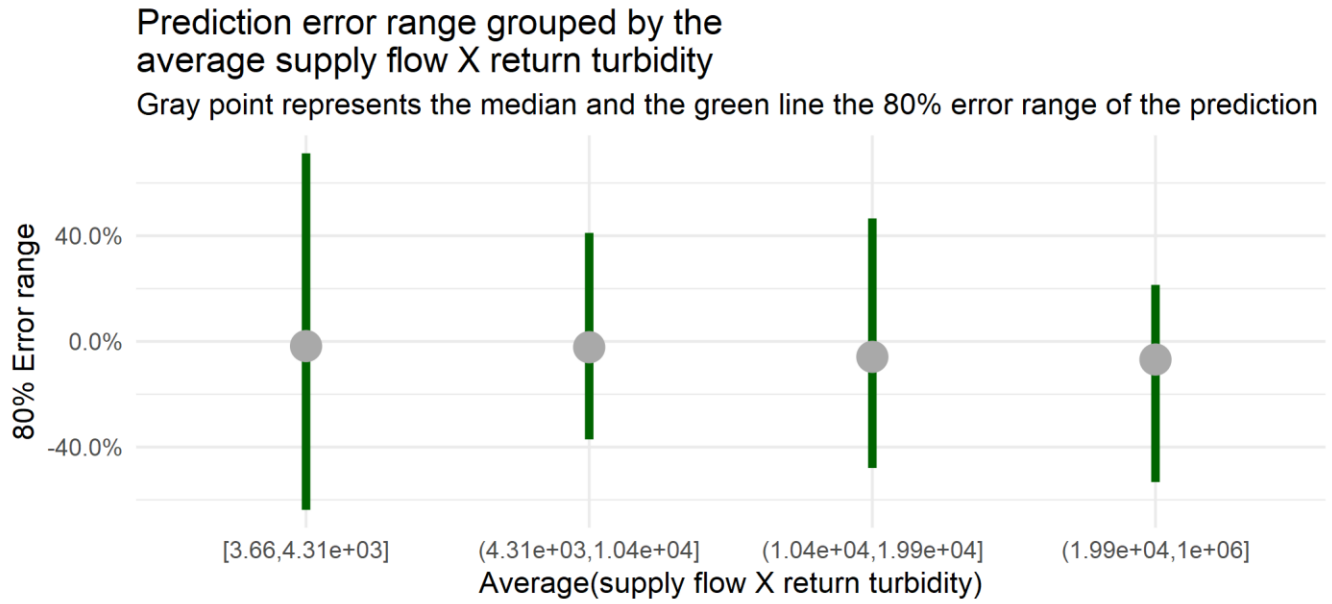
- The confidence level of the predictions increases when the supply flow median increases.



Confidence level

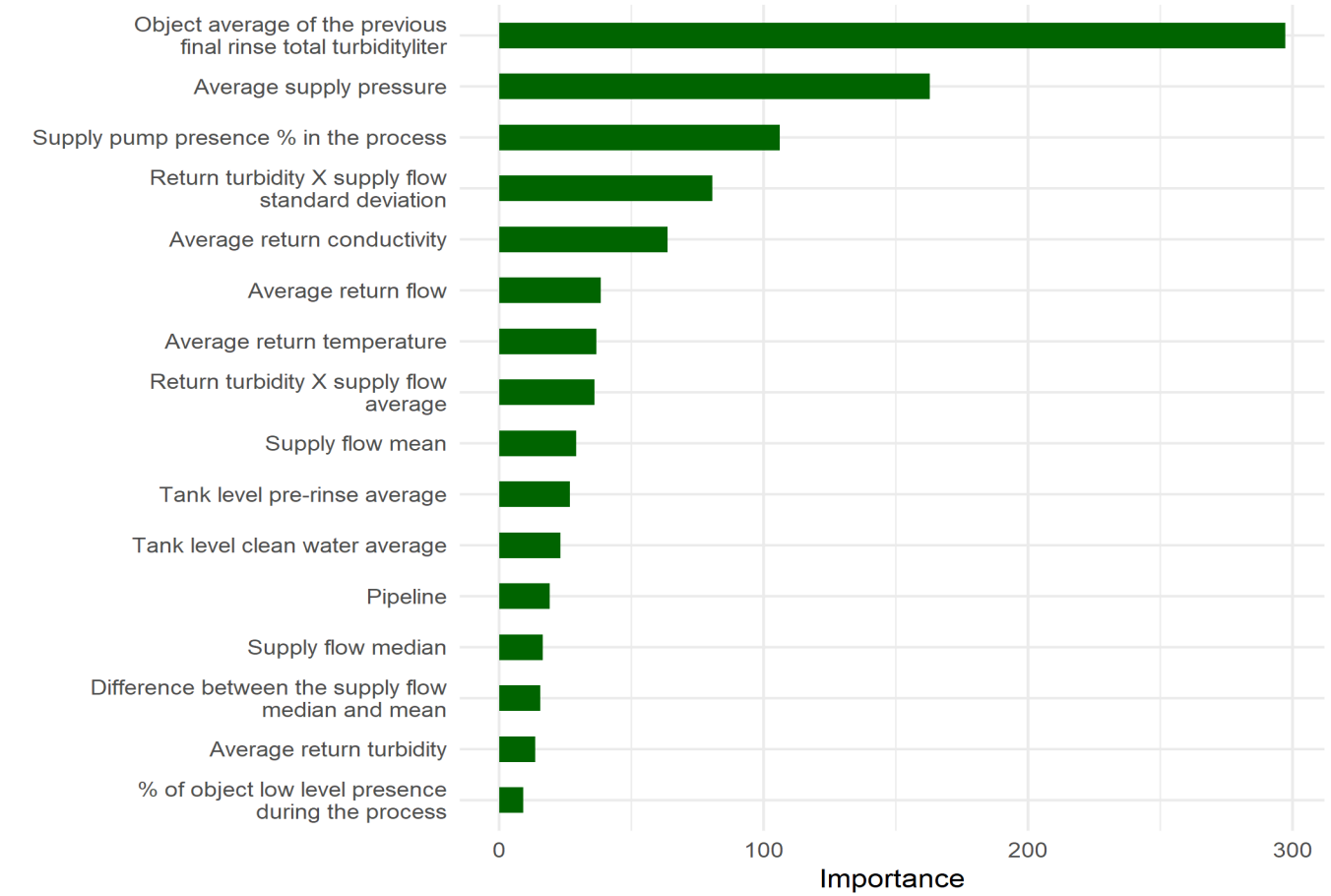
Supply flow and return turbidity

- One variable transformation that increased the accuracy of the model was the average of supply flow times the return turbidity. This variable showed as well an increase on the confidence prediction when the values were .



Variable importance

- Variable importance of all the variables used in our model, assuming the 4 phases previous to the final rinse phase were observed.



Conclusions

- Pipelines are a big factor in the final rinse total turbidity liter being L9 the one with the most turbidity produced among them.
- Tank level pre rinse average should be kept below 56
- Return temperature average when increased reduced considerably the final turbidity
- Acid phase is crucial to reduce the final rinse total turbidity liter for recipes with acid phase included
- Supply pressure average is the variable that impacted the most the final rinse total turbidity liter prediction