Table 1: Parameters manipulated

Training:testing ratio

Variables per split

(mtry)

Number of trees

(ntree)

| Parameter | Levels | Rationale |
|---------------------|----------------------|----------------------------------|
| Predictor variables | water quality, | Random forests only use a subset |
| | meteorology, both | of predictors for each tree |
| | | Different packages use different |
| Model package | randomForest, ranger | implementations and default |

parameters
The ratio can influence over-fitting

or under-fitting
The number of variables for each

split influence what data a tree

receives
Represents a trade-off between

model performance and

computational time

randomForest, ranger 0.7, 0.8, 0.9

2, 3, 4

100, 500, 1000