AQ Bench AOA Comparison

Marvin Ludwig

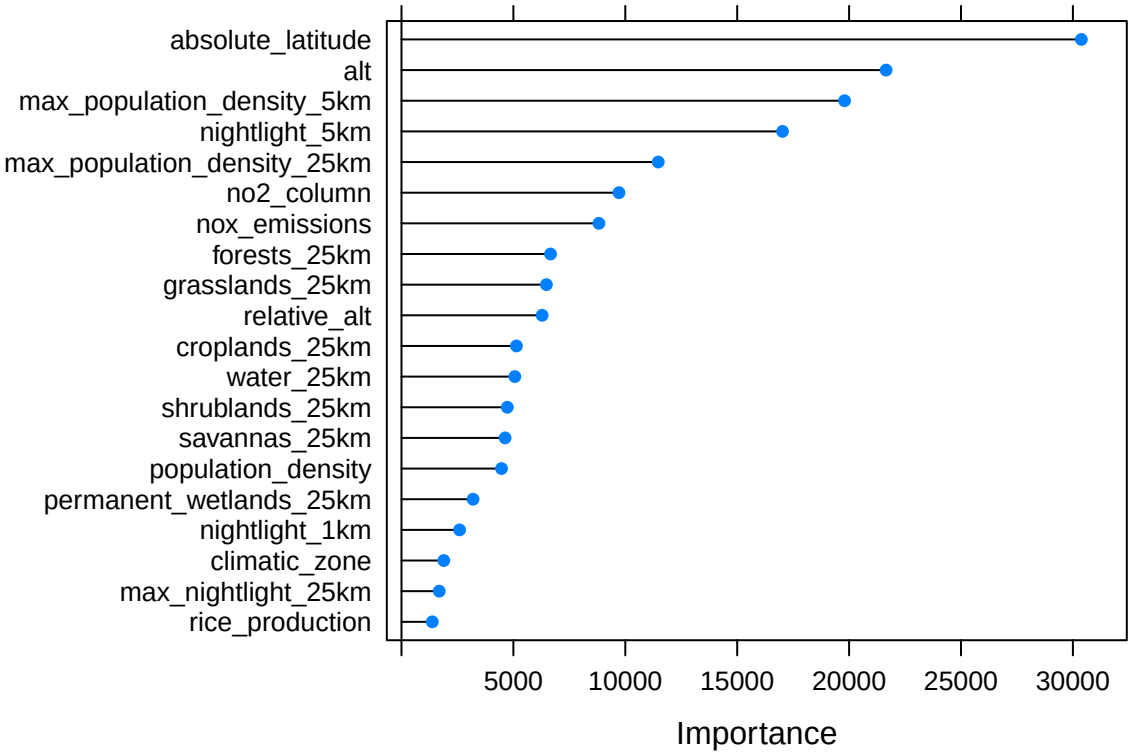
26/07/2021

All Predictors, CV-Juelich

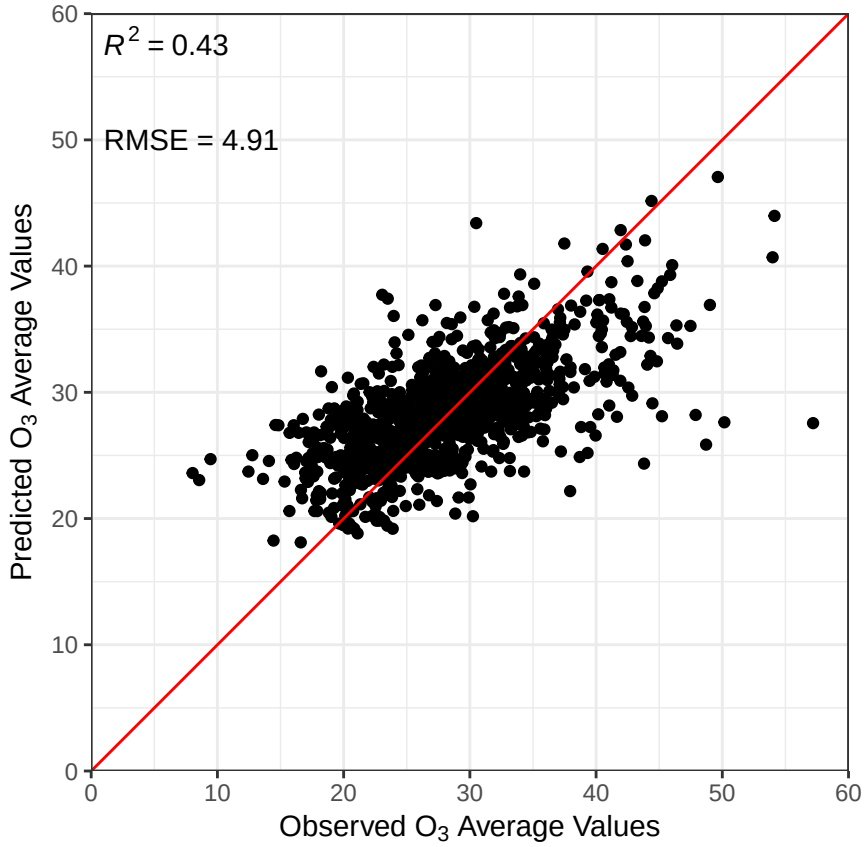
Model Infos

```
## Random Forest
##
## 4462 samples
##   20 predictor
##
## No pre-processing
## Resampling: Cross-Validated (4 fold)
## Summary of sample sizes: 3346, 3346, 3347, 3347
## Resampling results across tuning parameters:
##
##   mtry  RMSE      Rsquared   MAE
##    2    4.632545  0.4884201  3.387492
##    5    4.534638  0.4998935  3.312374
##    7    4.522061  0.5008224  3.296893
##   10    4.519086  0.4992388  3.294105
##   12    4.512786  0.5002000  3.289123
##   15    4.533746  0.4948408  3.305485
##   20    4.550528  0.4908901  3.311565
##
## Tuning parameter 'splitrule' was held constant at a value of variance
##
## Tuning parameter 'min.node.size' was held constant at a value of 5
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were mtry = 12, splitrule = variance
##  and min.node.size = 5.
```

Variable Importance (Gini)

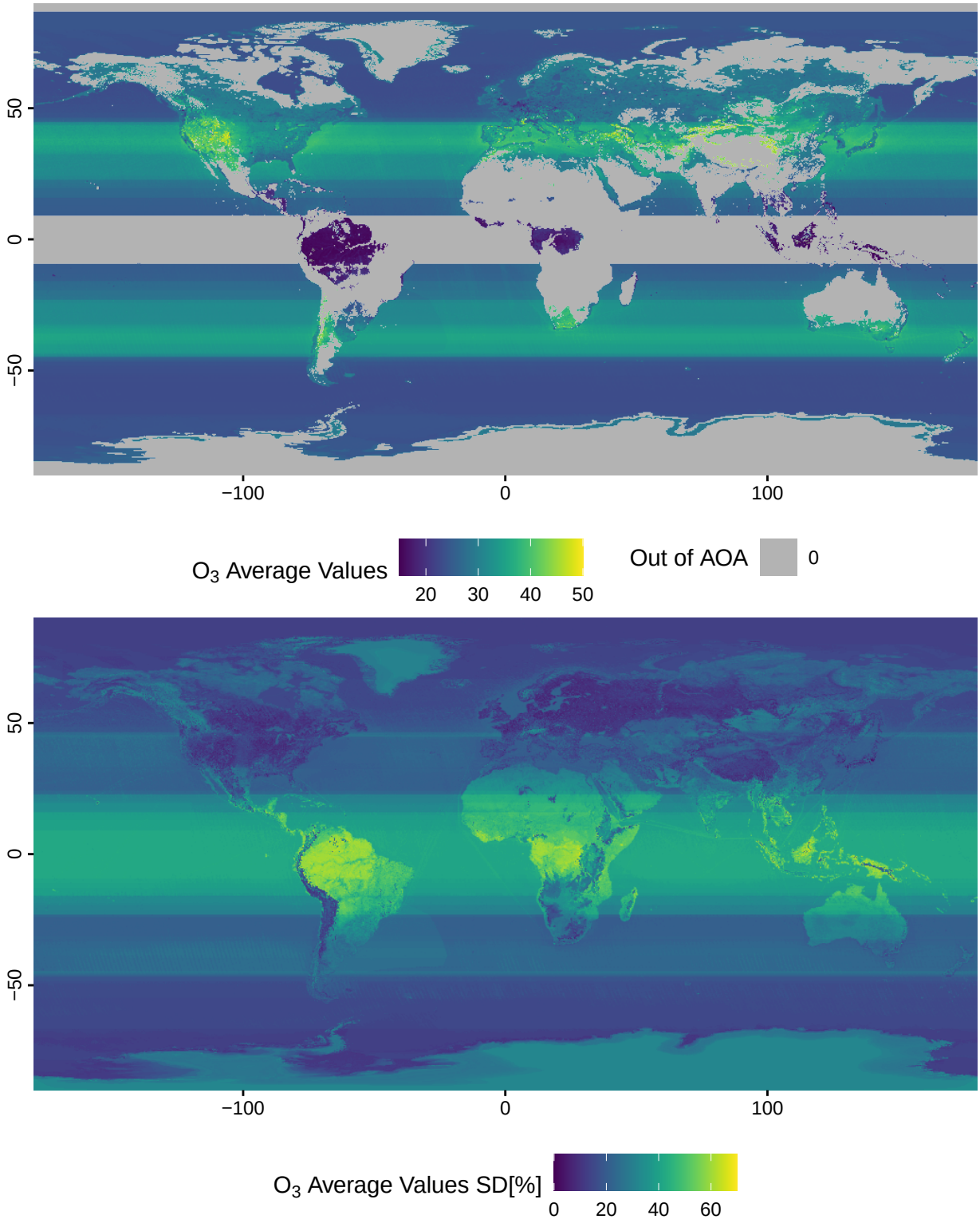


Validation on Test Set



AOA Stats

```
## [1] "Mean Distance in Training Data: 65020.55"
## [1] "DI threshold: 0.3289"
```



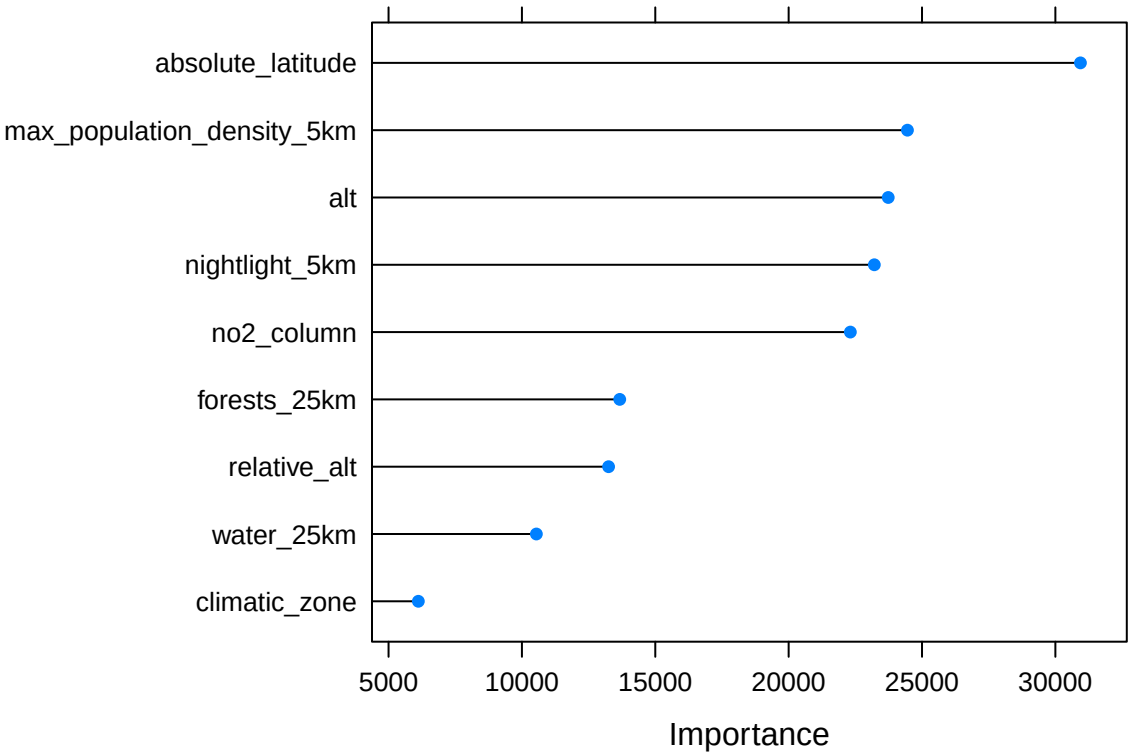
Forward Feature Selection, Juelich CV

Model Infos

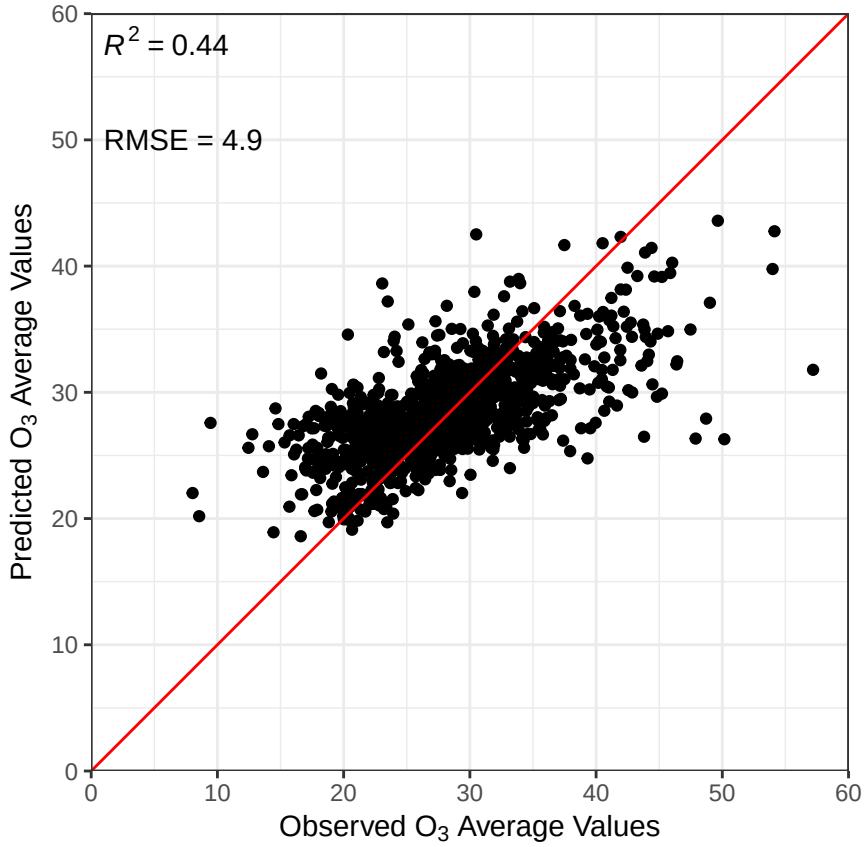
```
## Random Forest
```

```
##
## 4462 samples
##    9 predictor
##
## No pre-processing
## Resampling: Cross-Validated (4 fold)
## Summary of sample sizes: 3346, 3346, 3347, 3347
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   4.554247  0.4943451  3.321538
##
## Tuning parameter 'mtry' was held constant at a value of 2
## Tuning
##  parameter 'splitrule' was held constant at a value of variance
##
## Tuning parameter 'min.node.size' was held constant at a value of 5
```

Variable Importance (Gini)

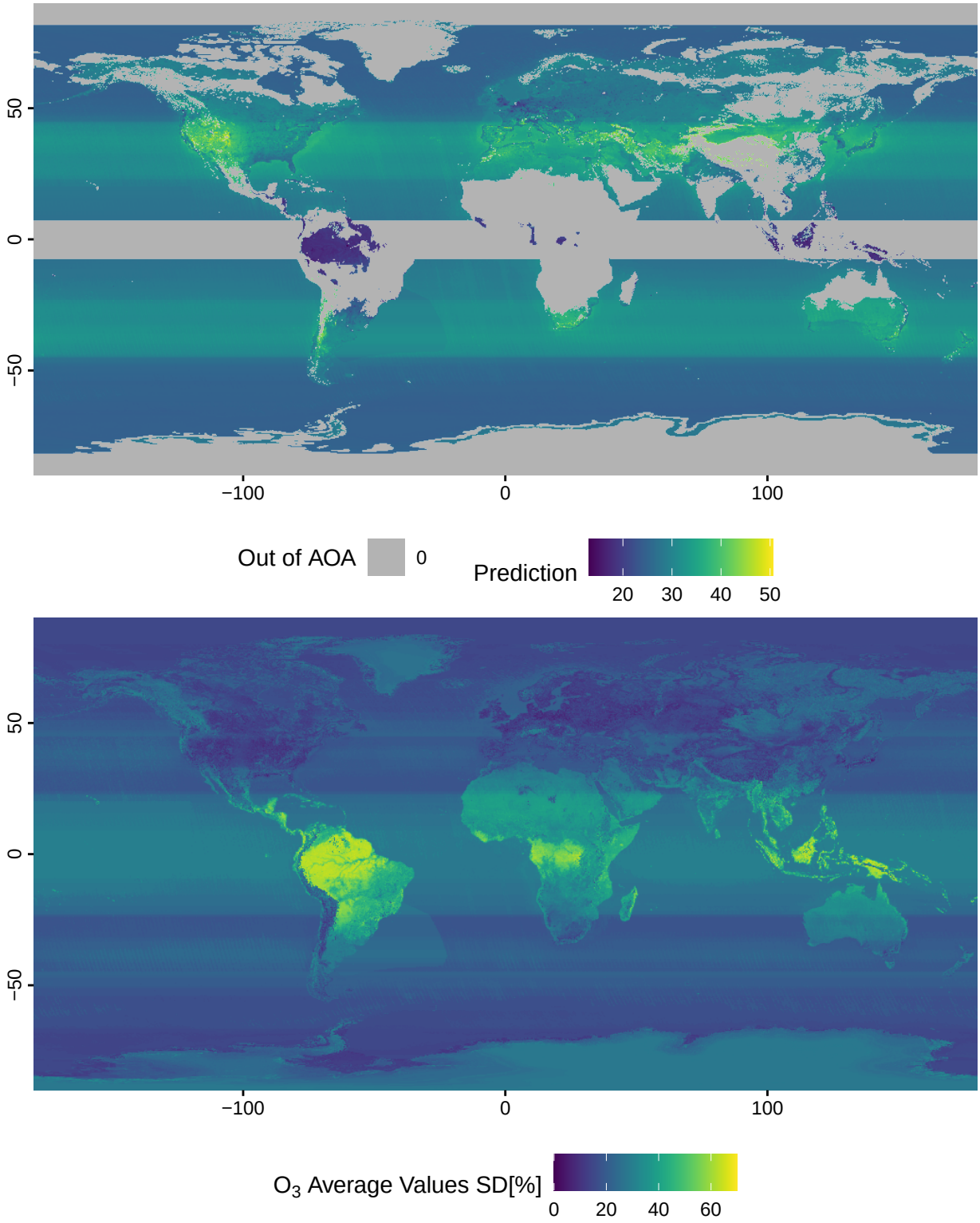


Validation on Test Set



AOA Stats

```
## [1] "Mean Distance in Training Data: 81344.13"
## [1] "DI threshold: 0.3206"
```



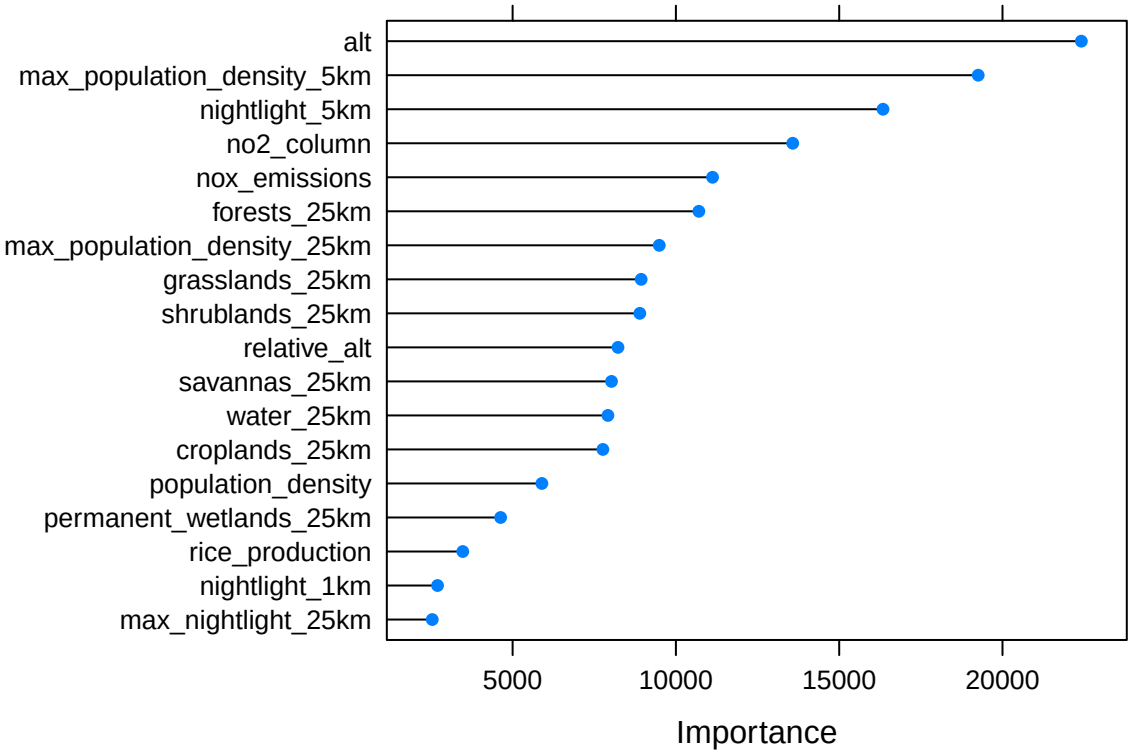
No Latitude and Climatic Zone, Juelich-CV

Model Infos

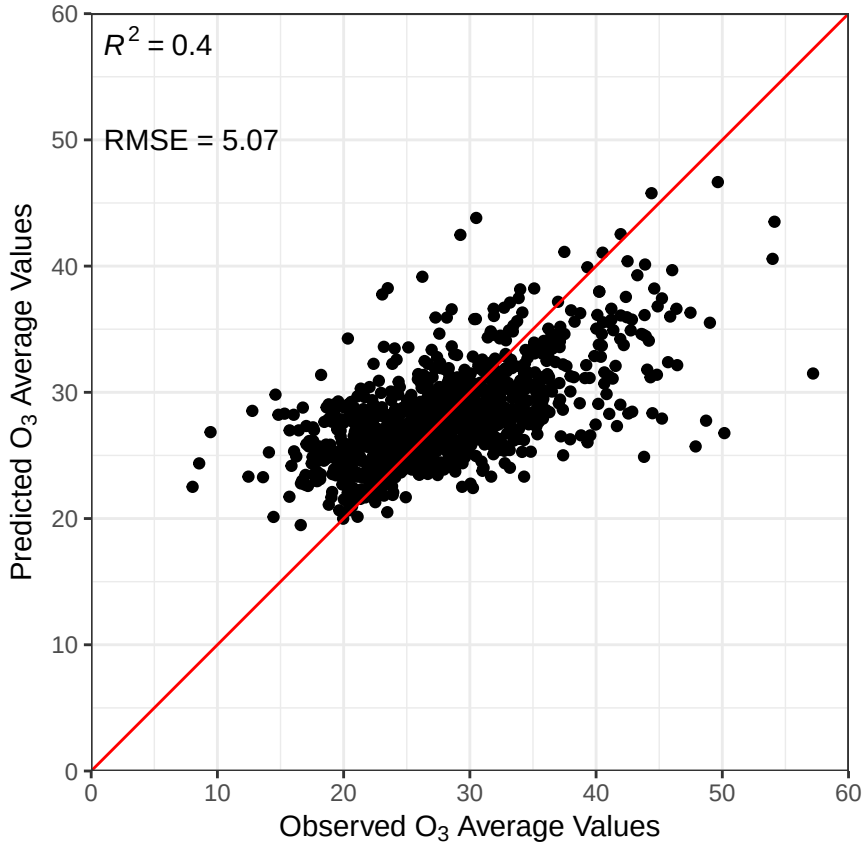
```
## Random Forest
```

```
##
## 4462 samples
##   18 predictor
##
## No pre-processing
## Resampling: Cross-Validated (4 fold)
## Summary of sample sizes: 3346, 3346, 3347, 3347
## Resampling results across tuning parameters:
##
##   mtry  RMSE      Rsquared   MAE
##    2    4.843144  0.4410863  3.596375
##    5    4.755628  0.4544175  3.533292
##    7    4.753775  0.4534554  3.529250
##   10    4.756186  0.4510575  3.525966
##   12    4.761877  0.4484919  3.523451
##   15    4.756936  0.4502775  3.530103
##
## Tuning parameter 'splitrule' was held constant at a value of variance
##
## Tuning parameter 'min.node.size' was held constant at a value of 5
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were mtry = 7, splitrule = variance
##  and min.node.size = 5.
```

Variable Importance (Gini)

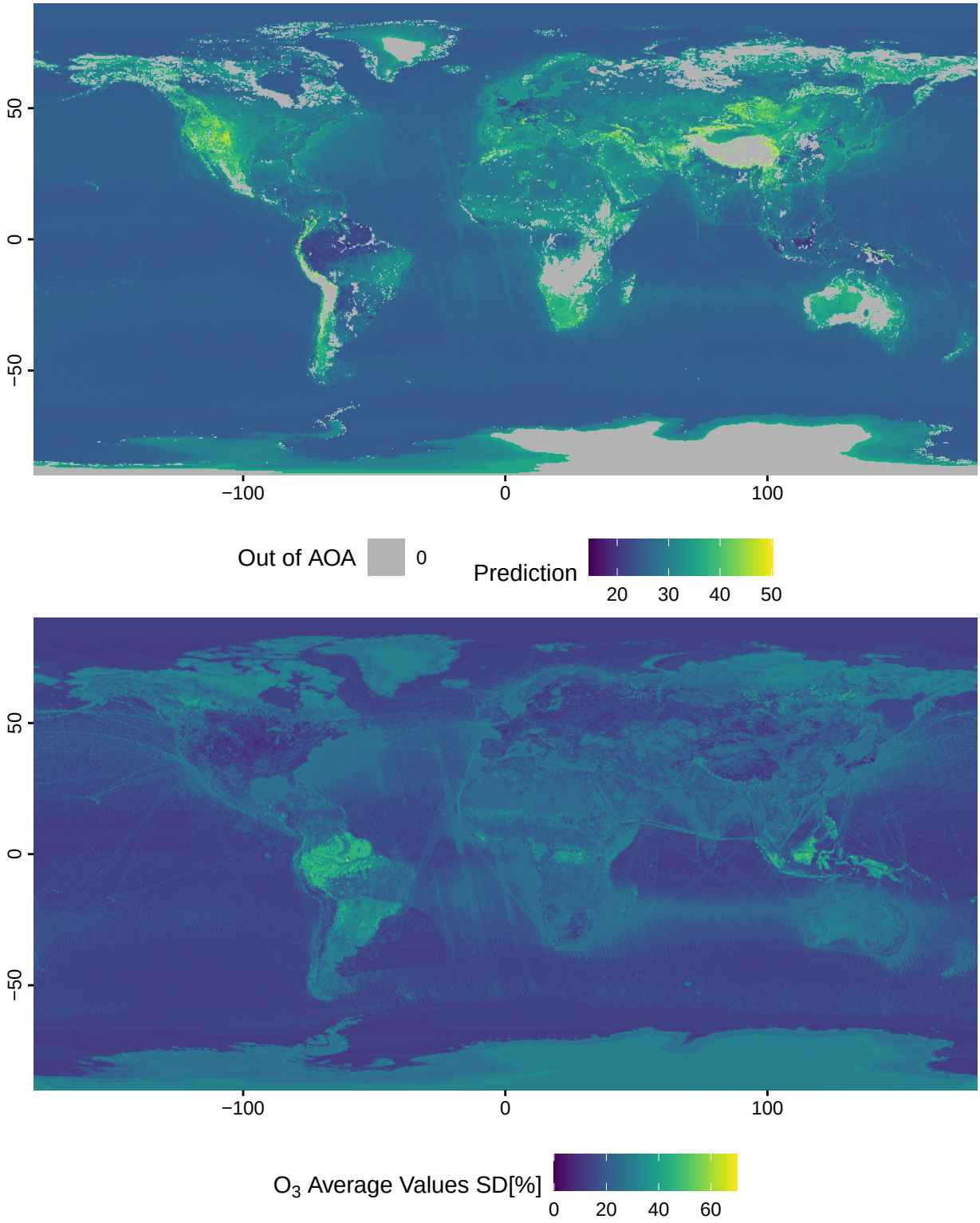


Validation on Test Set



AOA Stats

```
## [1] "Mean Distance in Training Data: 57528.2"
## [1] "DI threshold: 0.405"
```



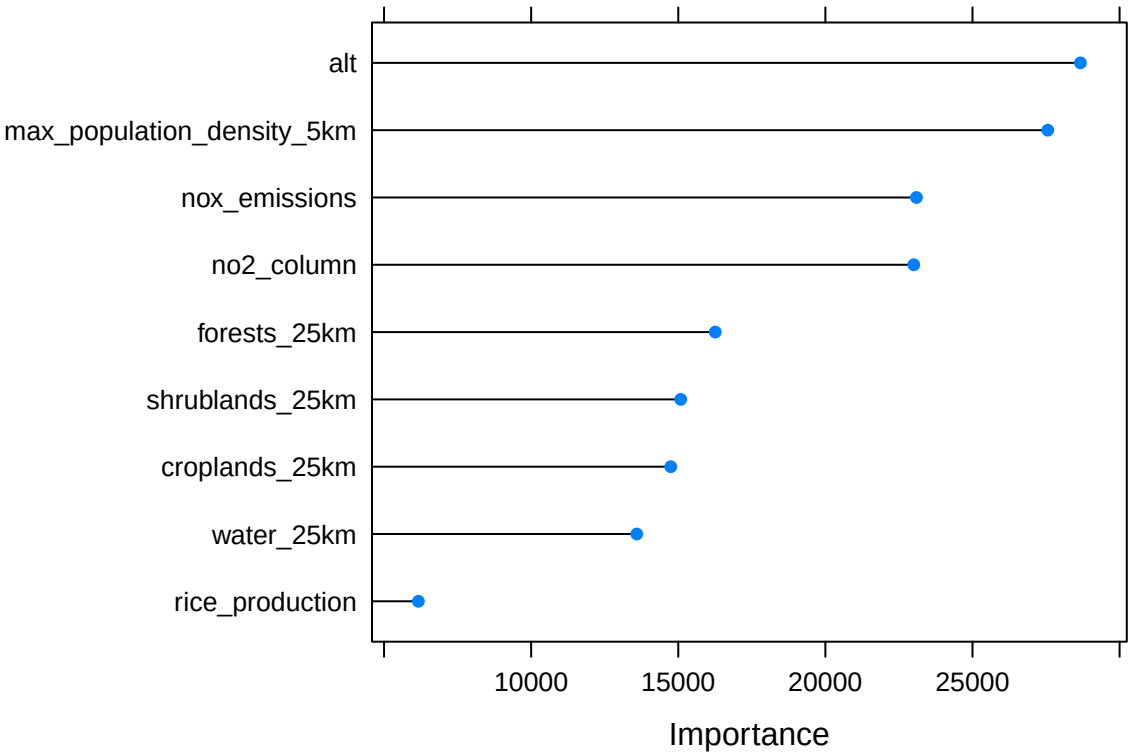
FFS, No Latitude and Climatic Zone, Juelich-CV

Model Infos

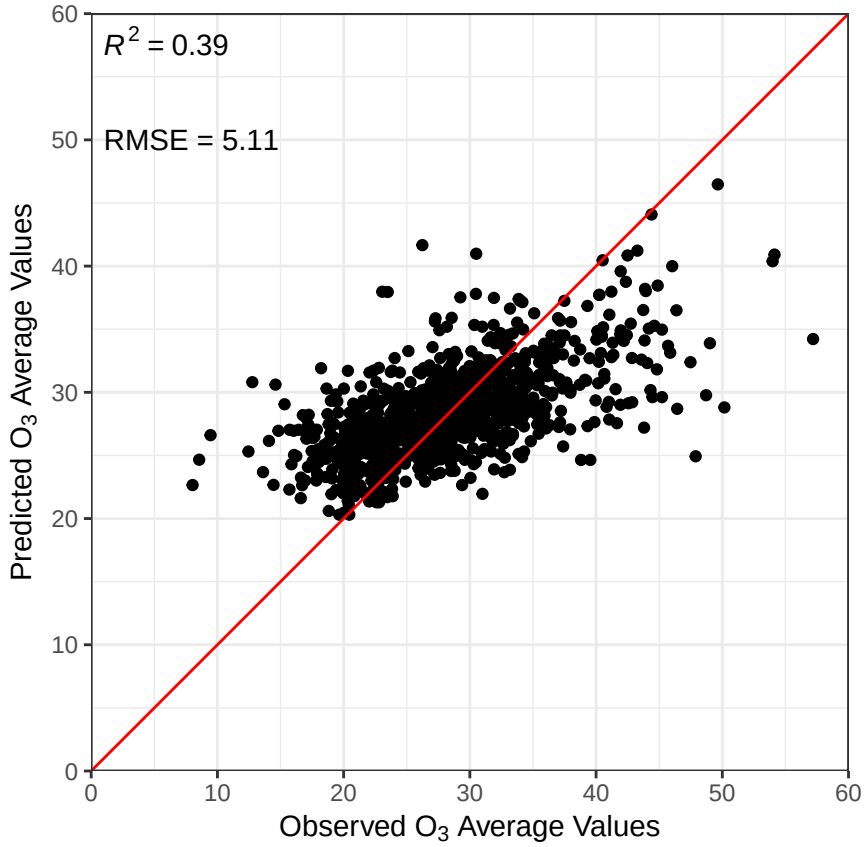
```
## Random Forest
```

```
##
## 4462 samples
##    9 predictor
##
## No pre-processing
## Resampling: Cross-Validated (4 fold)
## Summary of sample sizes: 3346, 3346, 3347, 3347
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   4.797561  0.4417623  3.568128
##
## Tuning parameter 'mtry' was held constant at a value of 2
## Tuning
##  parameter 'splitrule' was held constant at a value of variance
##
## Tuning parameter 'min.node.size' was held constant at a value of 5
```

Variable Importance (Gini)

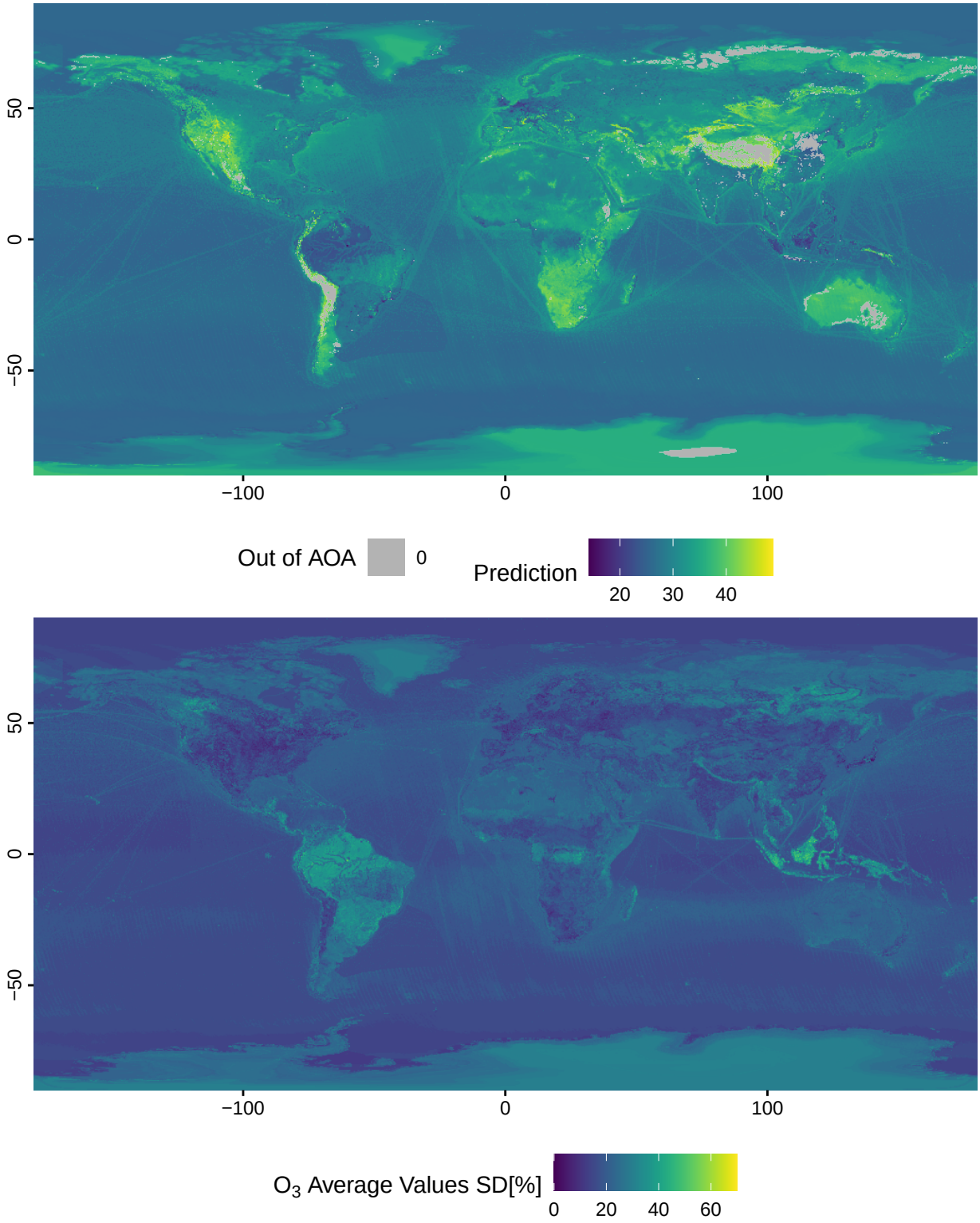


Validation on Test Set



AOA Stats

```
## [1] "Mean Distance in Training Data: 72049.12"
## [1] "DI threshold: 0.3214"
```


50
0
−50
−100
0
100
Out of AOA
0
Prediction
20
30
40
50
0
−50
−100
0
100
O3 Average Values SD[%]
0
20
40
60