**NAME: OYEDAYO OYELOWO STUDENT NUMBER:014717208 DATE: 19/10/2017**

Exercise 5 Tree variable imputation

1) Read “TreeVariableImputation.csv” into R

2) Predict mean diameter (d1.3), height and tree species for each tree using the ALS derived features.

Use k-NN method (R-package yaImpute). Get familiar with the following document Crookston

& Finley (2008) which is in Moodle.

• Test different explanatory variables (x)

• Test different values of k

• Test different strategies to select the nearest neighbours:

o Euclidean distance

o Most similar neighbour (MSN)

3) Analyze the results

• Calculate RMSE and bias

• Plot dependencies between measured and estimated attributes

Return pdf-document which includes your analysis of the achieved results. Analyze, what kind of dependencies were visible between the ALS derived features and field measured tree-level attributes. What were the best explanatory variables (ALS features) for mean diameter and height?

**Solution:**

Table showing the abbreviations used for the various methods.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NAMING | K | Method for searching kNN | Impute K | Impute method | Impute method factor |
| euc1 | 3 | euclidean | 3 | dstWeighted", | median |
| euc2 | 5 | euclidean | 5 | dstWeighted | median |
| euc3 | 3 | euclidean | 3 | mean | closest |
| euc4 | 5 | euclidean | 5 | mean | mean |
| msn1 | 3 | msn | 3 | dstWeighted | median |
| msn2 | 5 | msn | 5 | dstWeighted | median |
| msn3 | 3 | msn | 3 | mean | closest |
| msn4 | 5 | msn | 5 | mean | mean |
| rf | 3 | randomForest | 3 | mean | closest |

**BIAS DIAMETER**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | euc1 | euc2 | euc3 | euc4 | msn1 | msn2 | msn3 | msn4 | rf |
| diameter bias | 5.201636 | 3.602812 | 5.342949 | 0.225918 | 1.184429 | 1.184429 | 0.195513 | 1.344231 | 3.301282 |

**BIAS HEIGHT**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | euc1 | euc2 | euc3 | euc4 | msn1 | msn2 | msn3 | msn4 | rf |
| height bias | 0.30105 | 0.223005 | 0.312527 | 0.014044 | 0.053073 | 0.053073 | 0.01583 | 0.063252 | 0.171088 |

**CORRELATION**

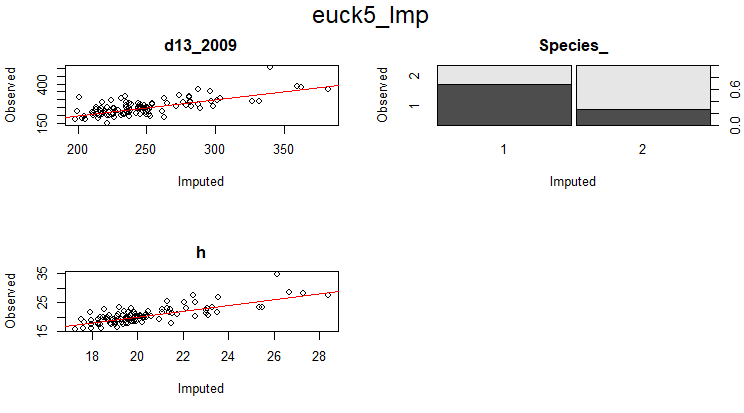
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | euc1 | euc2 | euc3 | euc4 | msn1 | msn2 | msn3 | msn4 | rf |
| d13\_2009 | 0.700007 | 0.714972 | 0.700615 | 0.709865 | 0.72576 | 0.742723 | 0.725377 | 0.739163 | 0.751047 |
| h | 0.752019 | 0.779727 | 0.756283 | 0.779215 | 0.783561 | 0.802102 | 0.784839 | 0.800756 | 0.794997 |
| Species2 | NA | NA | NA | NA | 0.433937 | 0.522126 | 0.436149 | 0.527491 | NA |

**RMSD**

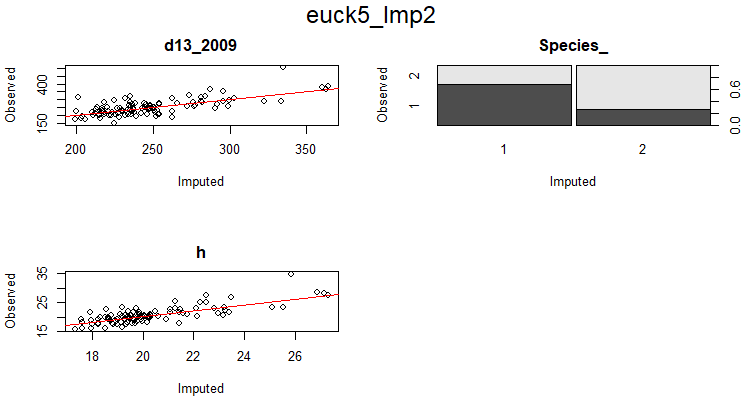
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | euc1 | euc2 | euc3 | euc4 | msn1 | msn2 | msn3 | msn4 | rf |
| d13\_2009 | 37.15376 | 36.15569 | 37.10373 | 36.46483 | 35.92286 | 34.47041 | 35.8647 | 34.66948 | 34.27145 |
| h | 1.964529 | 1.866156 | 1.9527 | 1.87528 | 1.848551 | 1.759918 | 1.839827 | 1.767138 | 1.798952 |
| Species2 | NA | NA | NA | NA | 0.474584 | 0.433132 | 0.473665 | 0.43101 | NA |

Root Mean Square difference is more preferable to correlation when trying to derive the level of accuracy of prediction (Warren 1971, Nicholas L. Crookston, Andrew O. Finley 2008. Therefore, I based my selection of the prediction method on that with the least error and the bias. From the results, the choice of K value affected the level of accuracy. When applying Euclidean and msn methods, a higher K value of 5 compared to 3, yielded more accurate predictions for both methods.

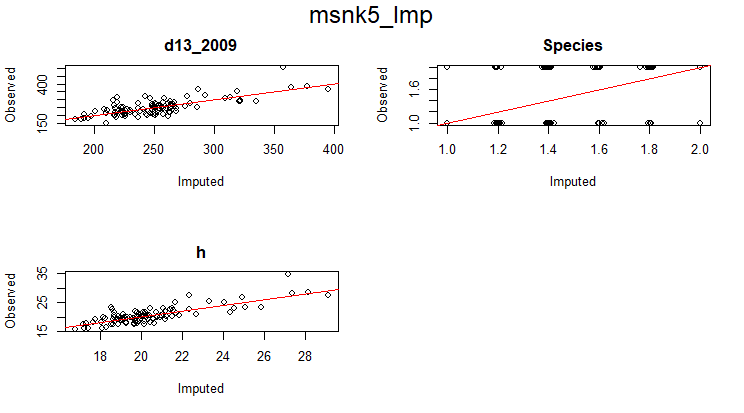
**euc2**



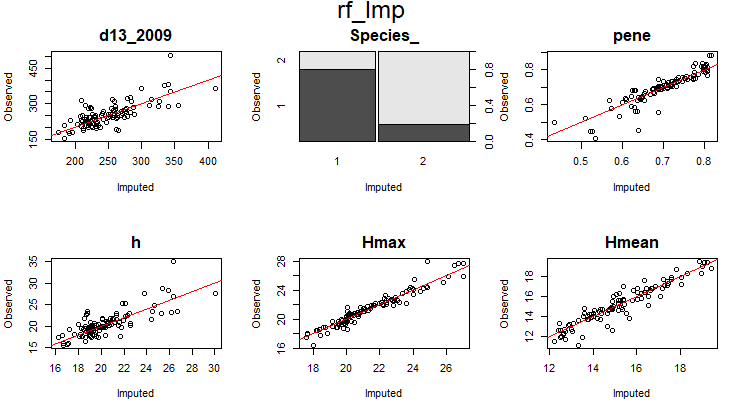
**euc4**



**msn4**



**rf**



The hmax , h90 and hmean appears to be the best explanatory variables. However, I chose the hmax and hmean, as all others seemed to be redundant. There is a deterministic dependency between the ALS features and the field data. The hmax was fair in predicting the height of trees and the diameter and also the species.

**Reference(s)**

Warren WG (1971). “Correlation or Regression: Bias or Precision.” Applied Statistics, 20(2), 148–164.

Nicholas L. Crookston, Andrew O. Finley, yaImpute: An R Package for kNN Imputation 2008 pg 4

Journal of Statistical Software