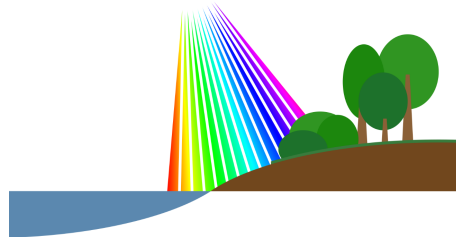


Zusammenfassung der Klassifikation: Vegetation Dominanzbestaende - Nonnenwerth - Sommeraufnahme 2019

mDRONES4rivers



Kontakt

mDRONES4rivers@bafg.de
Bundesanstalt für Gewässerkunde (BfG)
Am Mainzer Tor 1, 56068 Koblenz
www.bafg.de



Bearbeiter

Edvinas Rommel
Laura Giese
Frederik Kathoefer

Projektkoordination

Dr. Björn Baschek
Tel. 0261 / 1306 5395
baschek@bafg.de

Projektpartner

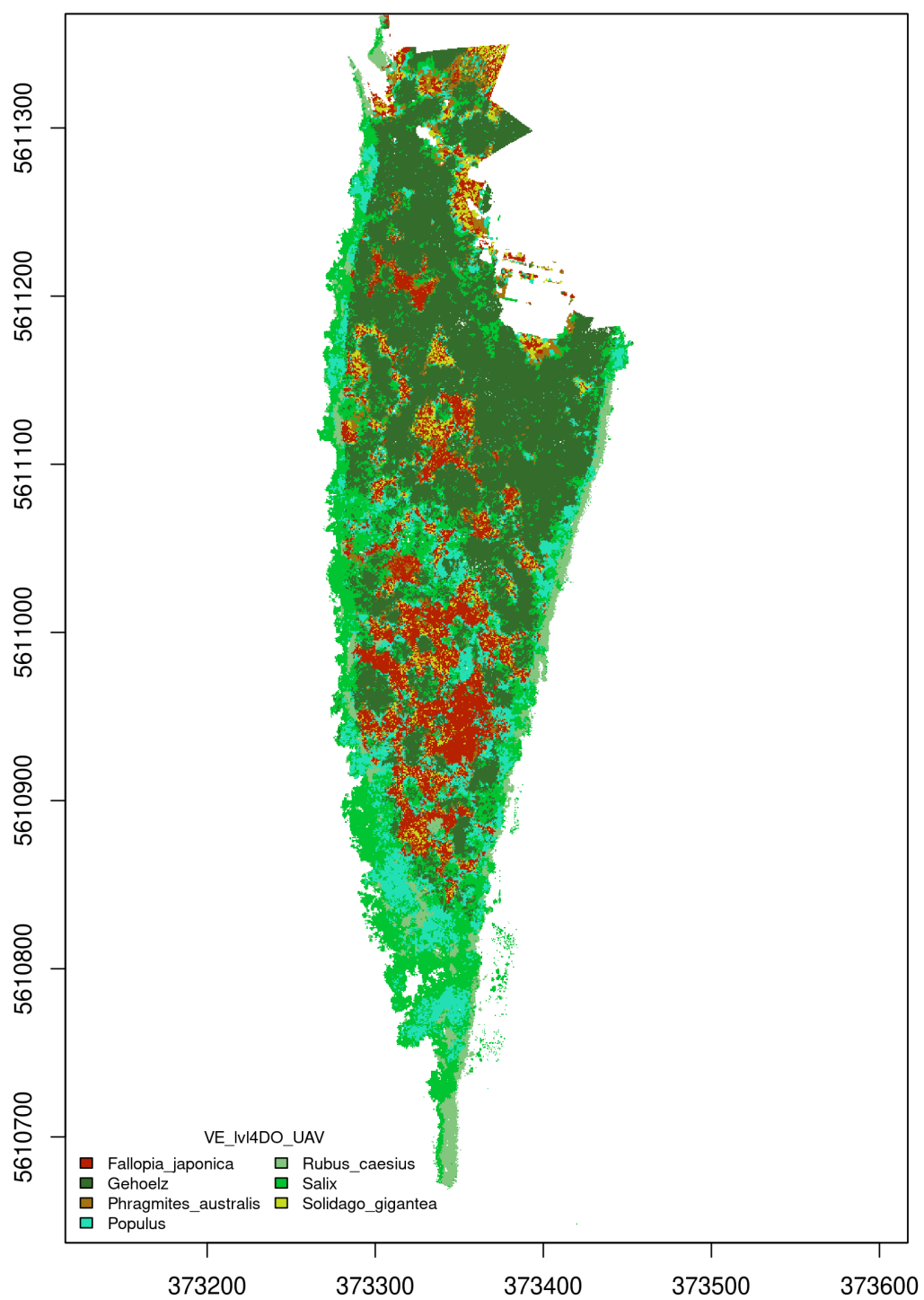


Verwendeter Algorithmus

Die Klassifikation wurde mit Random Forest (RF) durchgeführt

Übersichtskarte

DO_lvl4_NoW_2019_3 RF



Klassifikationsgüte

Die Klassifikationsgüte wird anhand einer Konfusionsmatrix und deren abgeleitete Gütemaße dargestellt. Die Zahlen in der Konfusionsmatrix sind die aufsummierten Werte von 500 Iterationen basierend auf dem Bootstrap Resampling Verfahren. Im Bootstrap Resampling werden die vorhandenen Referenzdaten zufällig in Trainings- und Validierungsdaten aufgeteilt.

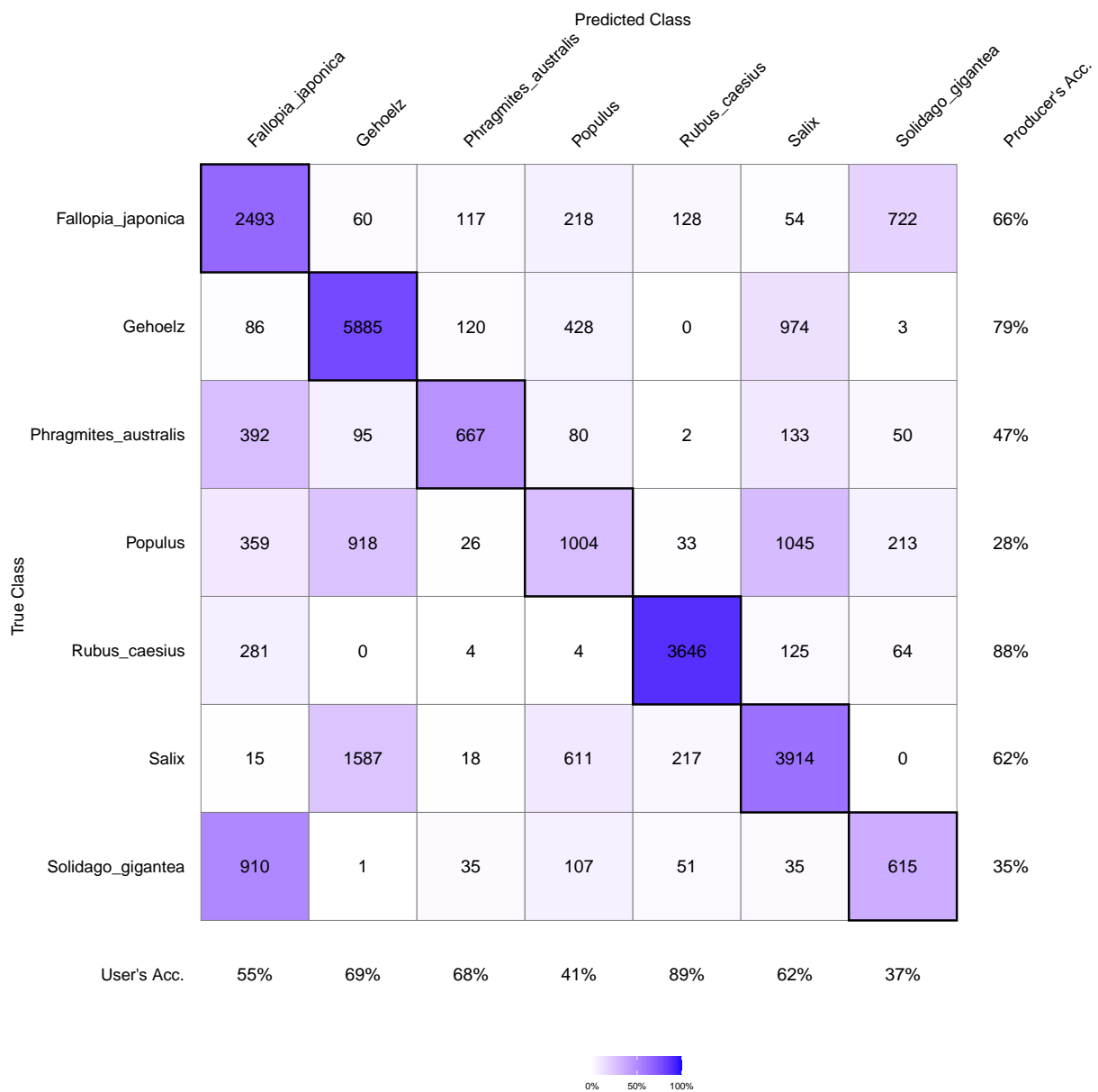
Auf Grundlage der Konfusionsmatrix werden die Producer-Accuracy (Herstellergenauigkeit), User-Accuracy (Nutzergenauigkeit), Overall-Accuracy (Gesamtklassifikationsgüte) und der Kappa Koeffizient berechnet

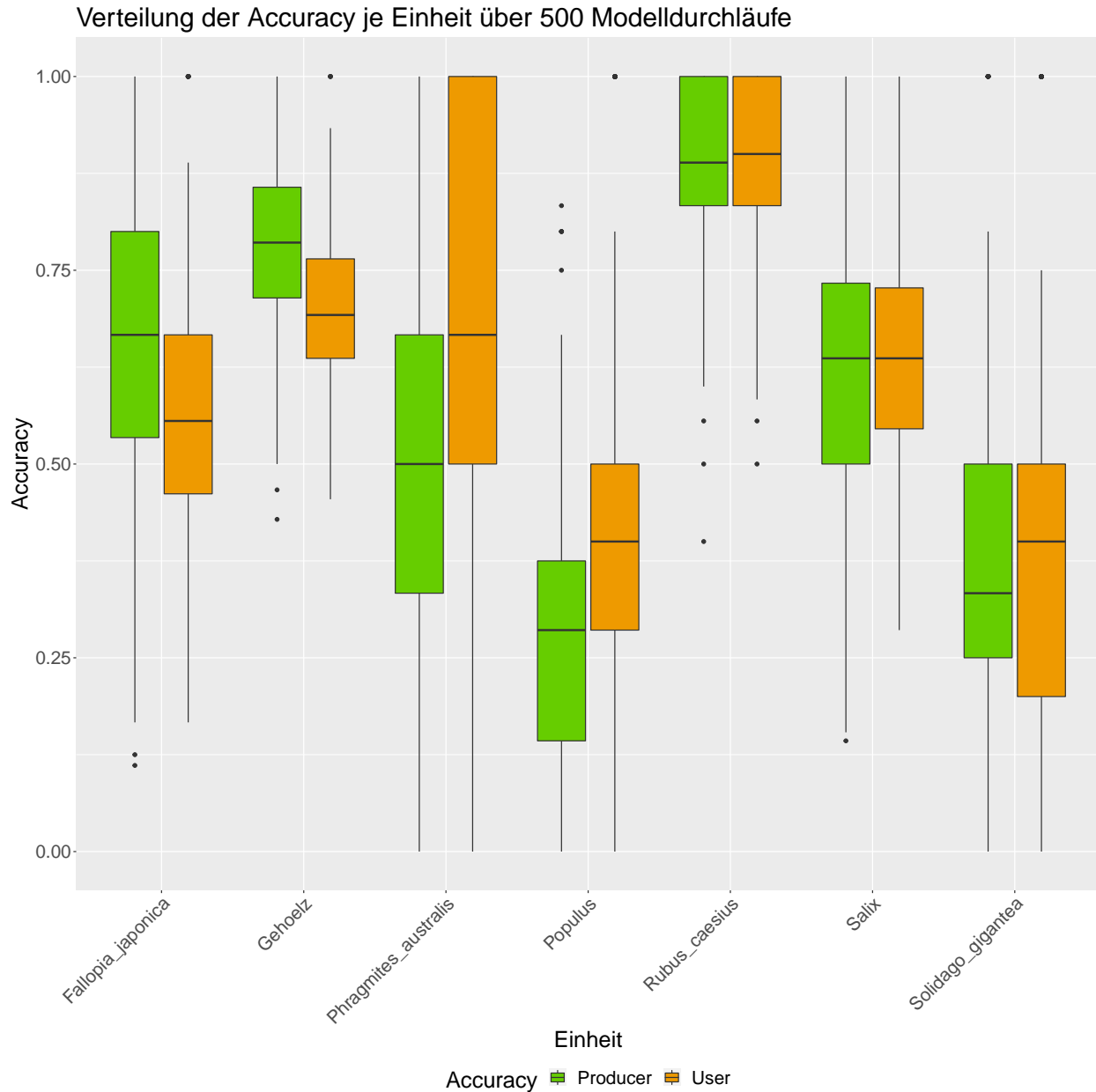
Random Forest (RF)

Overall Accuracy: **63.9 %**

Kappa: **0.56**

Confusion Matrix RF :





Variablen Selektion

Um die Gefahr einer Überanpassung des Models zu verringern und die Berechnungszeit zu verkürzen erfolgt eine Auswahl von wichtigen Variablen. Als Variablenselektionsmethode wurde die Algorithmunenabhängige Methode ranger-impurity aus dem R-Packet “ranger” gewählt. Ranger impurity ordnet die Variablen nach deren Wichtigkeit auf Basis der eines Entscheidungsbaumes indem die Knotenreinheit mit dem Gini Index berechnet wird. Je Höher der Wert, desto wichtiger die Variable.

Die optimale Anzahl an Variablen wird mit Hilfe einer “forward feature selection” bestimmt. Hierbei wird in einem iterativen Verfahren die Anzahl an Variablen schrittweise erhöht und ein Modell erstellt (angefangen bei der wichtigsten Variable). Es werden solange Variablen hinzugefügt bis eine Sättigung der Modelgüte erreicht ist.

Die wichtigsten Variablen sind im Nachfolgenden in absteigender Reihenfolge dargestellt.

Table 1: Ausgewählte Variablen für RF

Variable	Wert
mn_nDOM	11.99
mdiff_neig	7.149
mn_UFD_sd	6.893
sd_nDOM	6.656
mn_UFD_mn	6.539
sd_UFD_sd	5.598
sd_UFD_mn	5.363
mn_sr	4.758
mn_gvi	3.718
sd_sr	3.006
mn_ndwi	2.955
sd_ndwi	2.764
mn_ndre	2.724
sd_gvi	2.292
mn_NIR	2.233
sd_NIR	2.098
lewi	1.997
entr_NIR	1.985
mn_ndvi	1.899
sec_NIR	1.835
sd_ndre	1.818
corr_NIR	1.815
diss_NIR	1.811
mn_savi	1.798
cont_NIR	1.742
no_pix	1.733
mn_gndvi	1.706
hom_NIR	1.666
sd_ndvi	1.662
sd_totbr	1.605
sd_savi	1.571
sd_gndvi	1.515
mn_totbr	1.457

Anzahl Trainings- und Validierungsdaten

Verteilung der insgesamt **158** Referenzdaten über die Einheiten, die zum Trainieren und Validieren der Modelle genutzt werden:

Table 2: in situ Punkte: DO_lvl4_NoW_2019_3

Einheit	Anzahl
Fallopia_japonica	21
Gehoeiz	41
Phragmites_australis	8
Populus	20

Einheit	Anzahl
Rubus_caesius	23
Salix	35
Solidago_gigantea	10