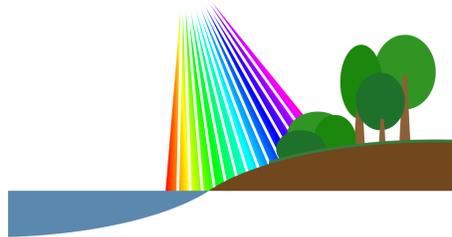


# Zusammenfassung der Klassifikation: Basisklassen - Nonnenwerth - Sommeraufnahme 2019

mDRONES4rivers



## Kontakt

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



## Bearbeiter

Edvinas Rommel  
Laura Giese  
Frederik Kathoefler

## Projektkoordination

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

## Projektpartner

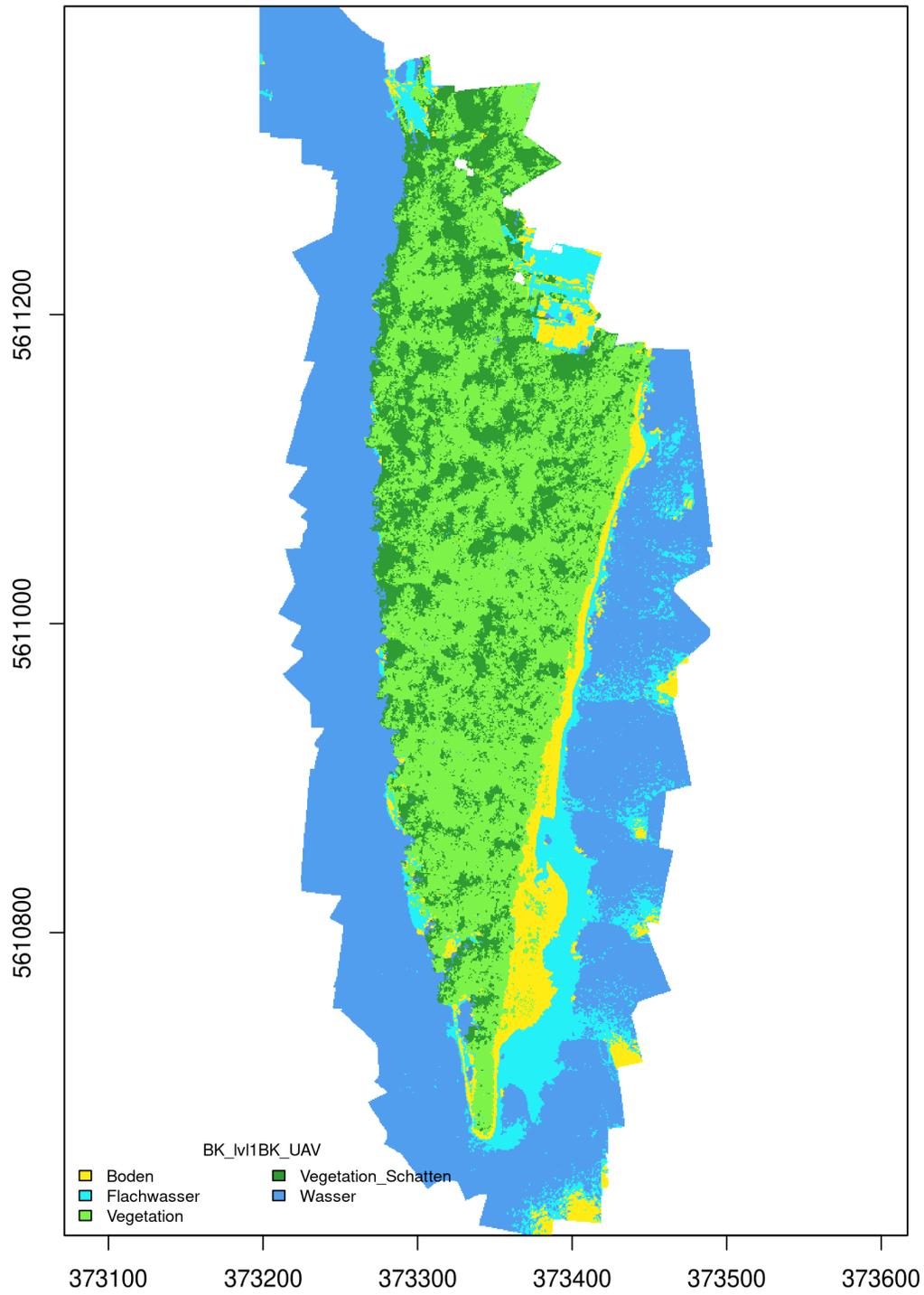


## **Verwendeter Algorithmus**

Die Klassifikation wurde mit Random Forest (RF) durchgeführt

## **Übersichtskarte**

BK\_lv11\_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

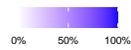
### **Random Forest (RF)**

Overall Accuracy: **91.4 %**

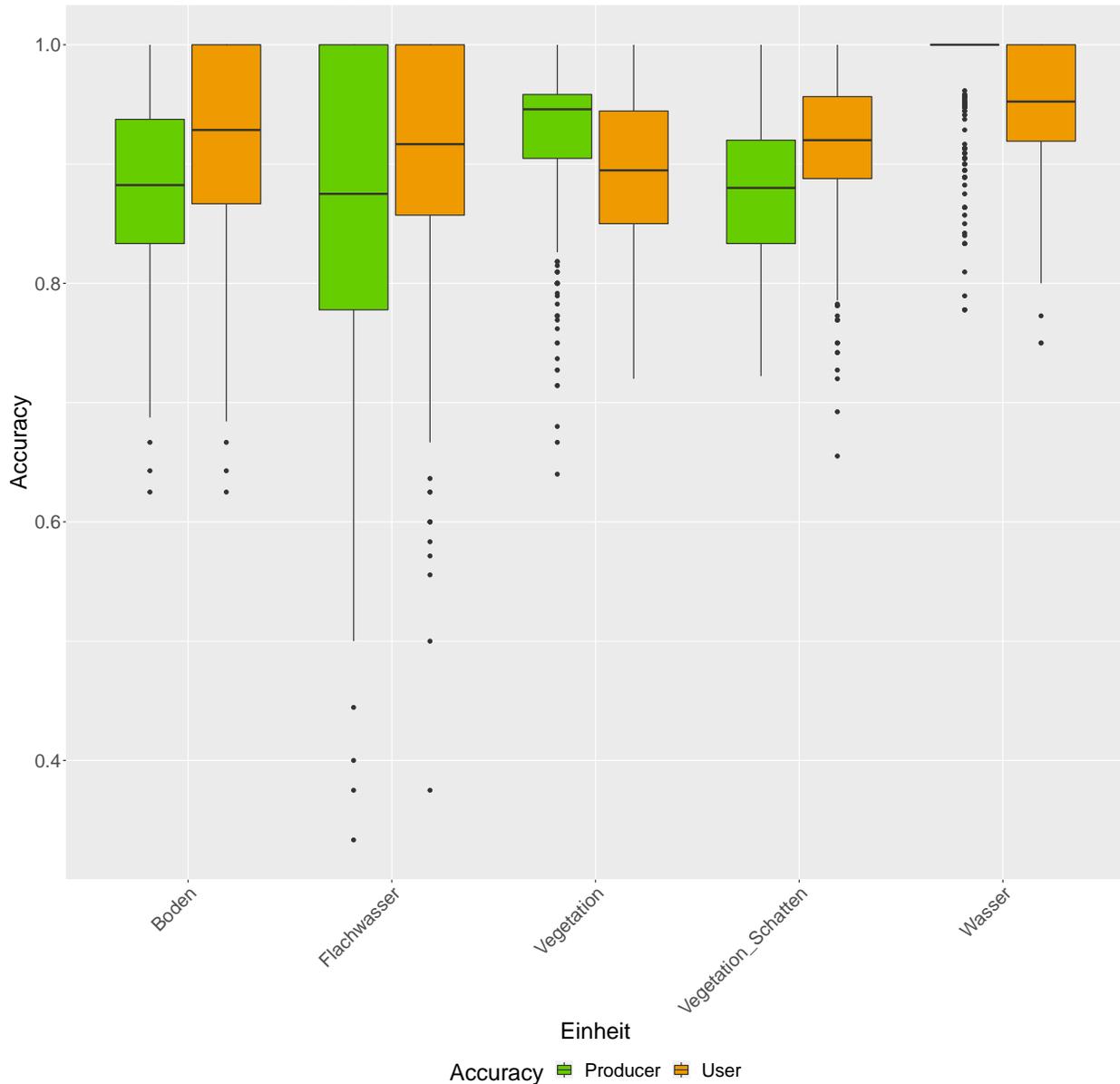
Kappa: **0.89**

Confusion Matrix RF :

		Predicted Class					Producer's Acc.
		Boden	Flachwasser	Vegetation	Vegetation_Schatten	Wasser	
True Class	Boden	6764	294	116	148	318	89%
	Flachwasser	502	3513	4	3	78	86%
	Vegetation	5	2	10303	823	2	93%
	Vegetation_Schatten	112	0	1143	10318	143	88%
	Wasser	81	55	5	8	9848	99%
User's Acc.		91%	91%	89%	91%	95%	



Verteilung der Accuracy je Einheit über 500 Modelldurchläufe



## Variablen Selektion

Um die Gefahr einer Überanpassung des Modells zu verringern und die Berechnungszeit zu verkürzen erfolgt eine Auswahl von wichtigen Variablen. Als Variablenselektionsmethode wurde die Algorithmunabhängige Methode ranger-impurity aus dem R-Paket "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_sr	9.22
sd_sr	8.393
mn_UFD_mn	8.34
mn_ndwi	7.971
mn_savi	7.676
sd_gvi	7.552
mn_gvi	7.318
mn_var_gr	6.958
mn_mean_gr	6.019
sd_ndwi	5.907
mn_ndre	5.571
mn_sec_gr	5.308
mn_gndvi	5.192
mn_entr_gr	4.935
mn_mean_kl	4.742
mn_totbr	4.534
sd_var_kl	4.438
mn_diss_gr	3.713
mn_ndvi	3.499
mn_UFD_sd	3.482
sd_savi	3.367
mn_hom_gr	2.907
mn_sec_kl	2.88

## Anzahl Trainings- und Validierungsdaten

Verteilung der insgesamt **245** Referenzdaten über die Einheiten, die zum Trainieren und Validieren der Modelle genutzt werden:

Table 2: in situ Punkte: BK\_lvl1\_NoW\_2019\_3

Einheit	Anzahl
Boden	42
Flachwasser	23
Vegetation	61
Vegetation_Schatten	64
Wasser	55