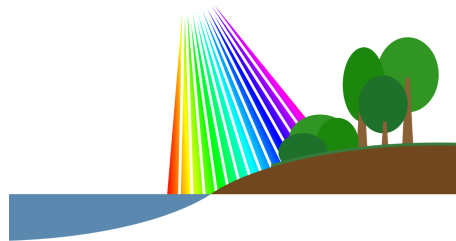


# Zusammenfassung der Klassifikation: Substrattypen - Nonnenwerth - Sommeraufnahme 2019

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

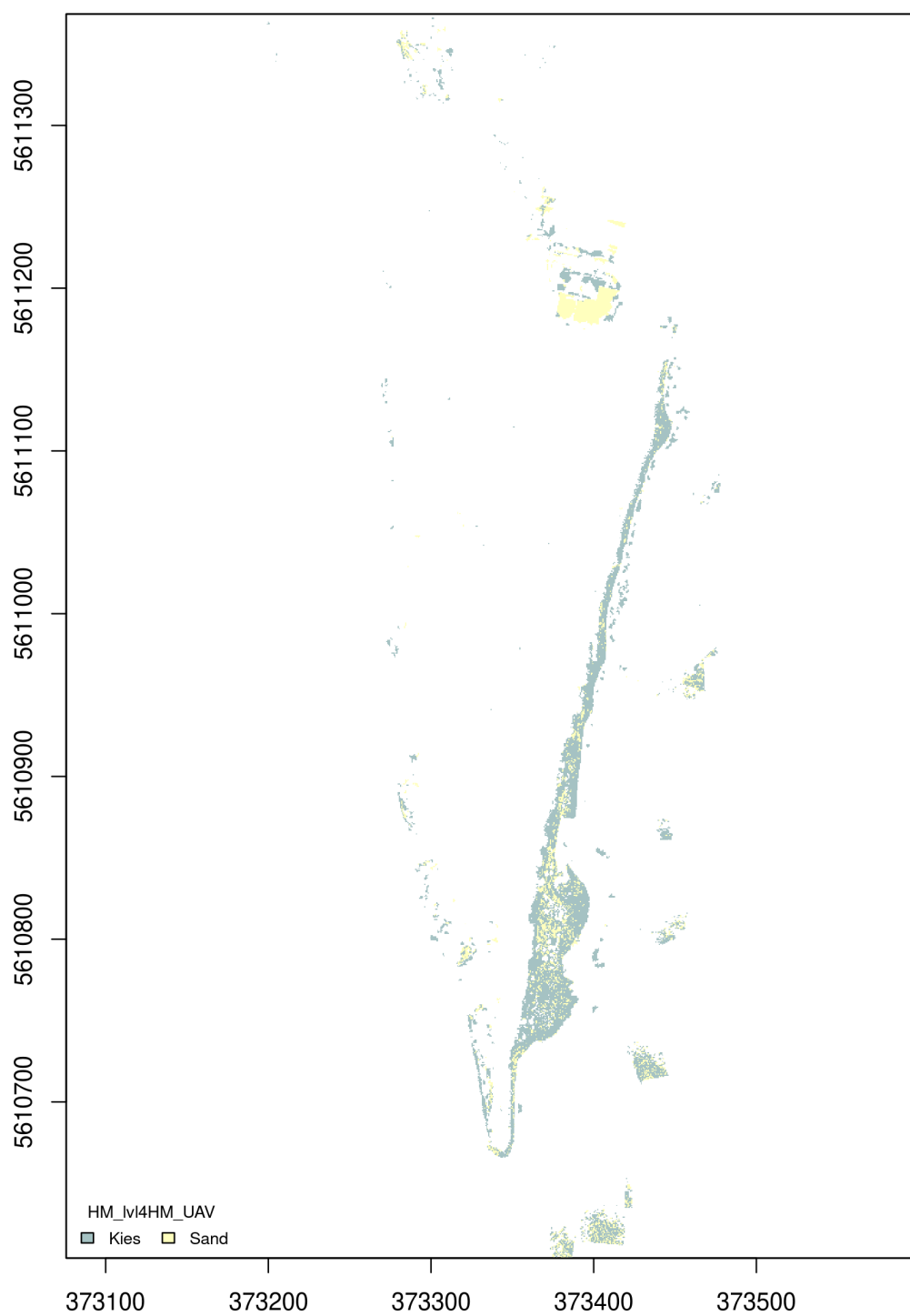


## Verwendeter Algorithmus

Die Klassifikation wurde mit Random Forest (RF) durchgeführt

## Übersichtskarte

HM\_lv14\_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.

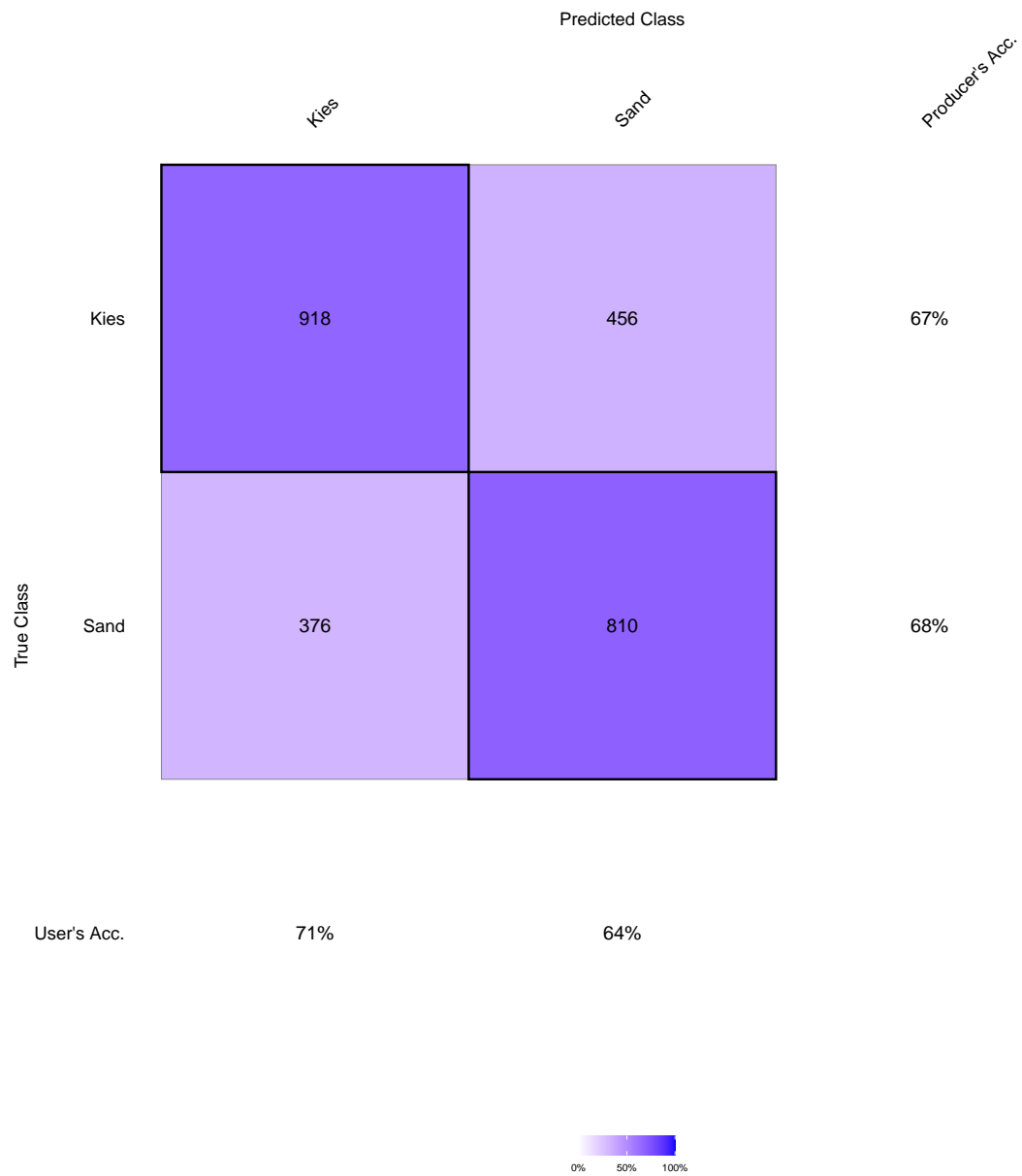
Auf Grundlage der Konfusionsmatrix werden die Producer-Accuracy (Herstellergenauigkeit), User-Accuracy (Nutzergenauigkeit), Overall-Accuracy (Gesamtklassifikationsgüte) und der Kappa Koeffizient berechnet

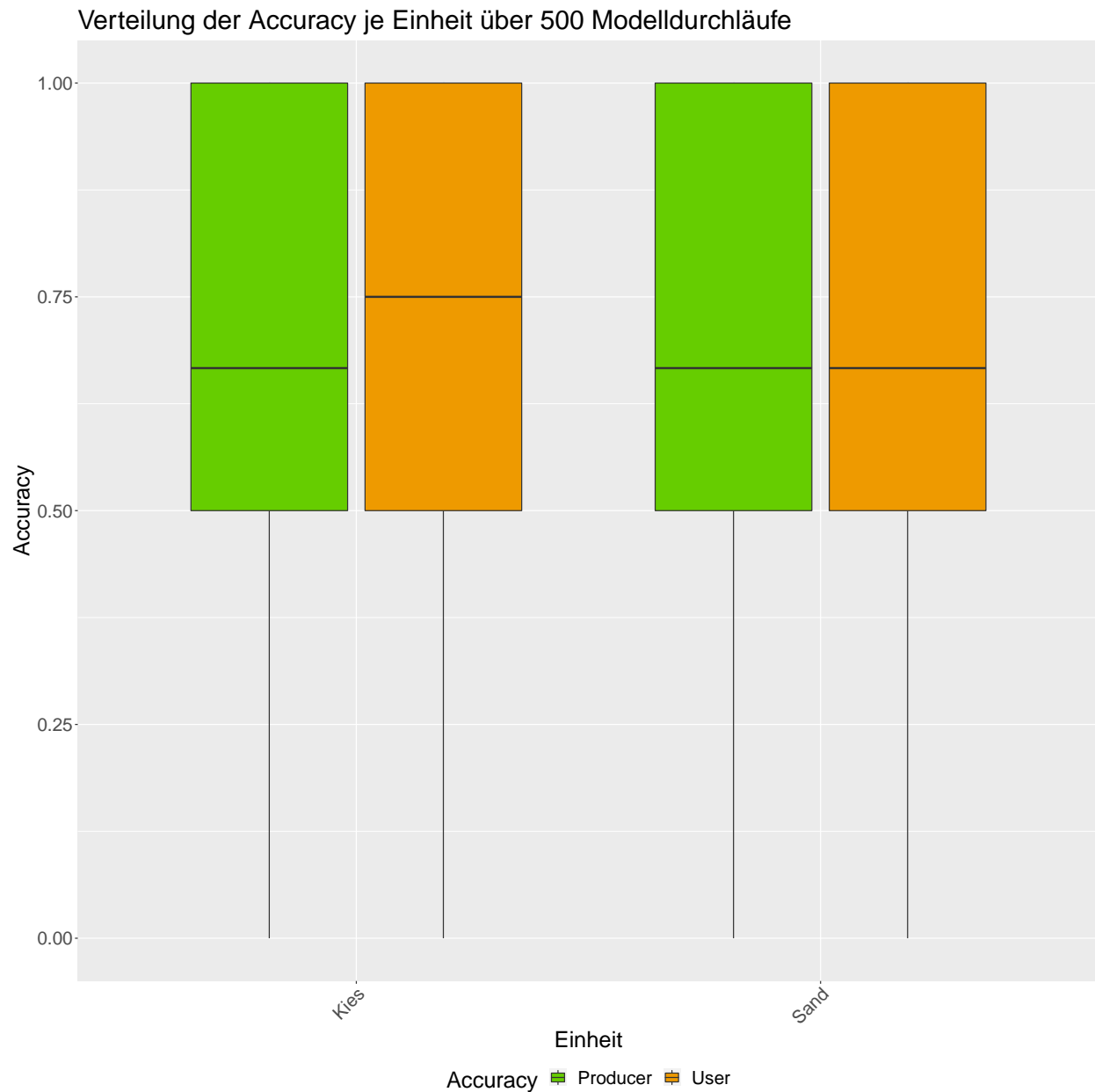
### **Random Forest (RF)**

Overall Accuracy: **67.8 %**

Kappa: **0.35**

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 Algorithmenunabhängige 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
diss_NIR	0.3022
sd_sec_kl	0.2979
mn_mean_kl	0.2955
mn_UFD_sd	0.2914
sd_entr_kl	0.2814

## Anzahl Trainings- und Validierungsdaten

Verteilung der insgesamt **15** Referenzdaten über die Einheiten, die zum Trainieren und Validieren der Modelle genutzt werden:

Table 2: in situ Punkte: HM\_lvl4\_NoW\_2019\_3

Einheit	Anzahl
Kies	8
Sand	7