

Area of Applicability - Practise

Hendrik Lüning und Maximilian Elfers



Usage / Workflow

- Funktion **Area of Applicability** (AOA) berechnet / schätzt
 - **Dissimilarity Index (DI)**
 - **Abgeleitete AOA von räumlichen Vorhersage Modellen**
 - Hierfür wird er Abstand neuer Daten im Raum zu den für das Modelltraining verwendeten Daten berücksichtigt

 - Daten ähnlich zu bekannten -> kurze Distanz (DI tendiert zu 0)
 - Daten nicht ähnlich zu bekannten -> hohe Distanz (DI höher)
 - **Merke: Je höher der DI-Wert, desto schlechter die räumliche Vorhersage in dem Bereich**
-

Standard Ablauf

Ein Standard Ablauf zum verwenden der Funktion könnte so aussehen:

- Beispieldaten vorbereiten
 - (Beispieldaten visualisieren)
 - Modell trainieren
 - AOA des trainierten Modells berechnen
-
- (AOA kann auch ohne ein trainiertes Modell berechnet werden)

aoa(**newdata**, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)

Ein **RasterStack**, **RasterBrick**, stars object, SpatRaster oder data.frame
Beinhaltet die **Daten**, für die eine **Vorhersage** gemacht werden soll.

`aoa(newdata, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)`

Ein mit **caret** erstelltes **train-Objekt**, aus dem Gewichte (basierend auf der **Wichtigkeit der Variablen**) sowie **cross-validation folds** extrahiert werden.

`aoa(newdata, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)`

Ein **trainDI-Objekt**. **Optional**, wenn [trainID](#) berechnet wurde.

`aoa(newdata, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)`

Ein **Clusterobjekt**, das z. B. mit **doParallel** erstellt wurde. **Optional**. Sollte nur verwendet werden, wenn **newdata** groß ist.

`aoa(newdata, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)`

Ein **data.frame**, der die für das Modelltraining **verwendeten Daten** enthält. **Optional**. Nur erforderlich, wenn kein Modell angegeben wird.

`aoa(newdata, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)`

Ein **data.frame** mit **Gewichten** für jede Variable. **Optional**. Nur erforderlich, wenn kein Modell angegeben wird.

`aoa(newdata, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)`

character vector der **erklärenden Variablen**. Wenn "all", dann werden alle Variablen des Modells verwendet, oder wenn kein Modell angegeben ist, dann aus dem Trainingsdatensatz.

`aoa(newdata, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)`

Liste oder **Vektor**. Entweder eine Liste, in der jedes Element die Datenpunkte enthält, die während der Kreuzvalidierungsiteration zum Testen verwendet wurden (d. h. zurückgehaltene Daten). Oder ein Vektor, der die ID der Falte für jeden Trainingspunkt enthält. Nur erforderlich, wenn kein Modell angegeben wird.

`aoa(newdata, model, trainDI, cl, train, weight, variables, CVtest, CVtrain)`

Liste. Jedes Element enthält die **Datenpunkte**, die während der **Kreuzvalidierungsiteration** für das Training verwendet wurden (d. h. zurückgehaltene Daten). Nur erforderlich, wenn kein Modell angegeben ist, und nur erforderlich, wenn CVtrain nicht das Gegenteil von CVtest ist (d. h., wenn ein Datenpunkt nicht zum Testen verwendet wird, wird er zum Training verwendet). Relevant, wenn einige Datenpunkte ausgeschlossen werden, z. B. bei Verwendung von [nndm](#).

Beispiel

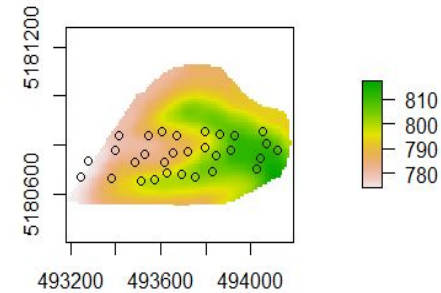
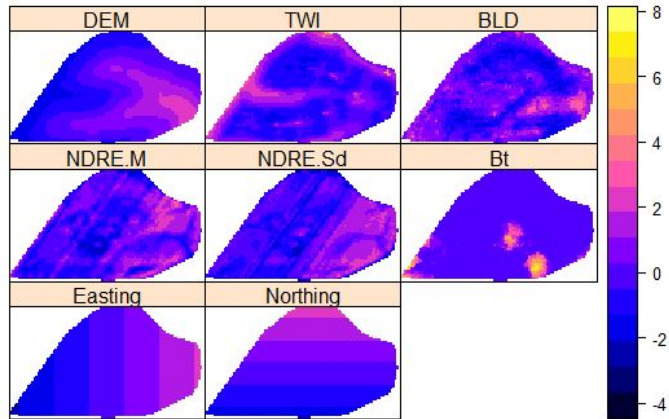
Beispieldaten vorbereiten

```
8 dat <- regm.VW
9 dat <- aggregate(dat[,c("VW", "Easting", "Northing")],
10                 by=list(as.character(dat$SOURCEID)), mean)
11 pts <- st_as_sf(dat, coords=c("Easting", "Northing"))
12 pts$ID <- 1:nrow(pts)
13 set.seed(100)
14 pts <- pts[1:30,]
15 studyArea <- stack("Dateinpfad/predictors_2012-03-25.grd")
16 trainDat <- extract(studyArea, pts, df=TRUE)
17 trainDat <- merge(trainDat, pts, by.x="ID", by.y="ID")
```

(Beispieldaten visualisieren)

```
20  spplot(scale(studyArea))  
21  plot(studyArea$DEM)  
22  plot(pts[,1],add=TRUE,col="black")
```

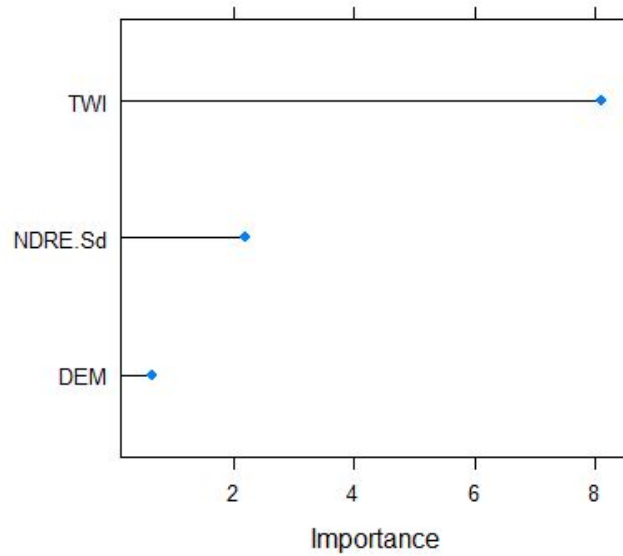
(Beispieldaten visualisieren) - Plots



Modell trainieren

```
25 set.seed(100)
26 variables <- c("DEM", "NDRE.Sd", "TWI")
27 model <- train(trainDat[,which(names(trainDat)%in%variables)],
28               trainDat$Vw, method="rf", importance=TRUE, tuneLength=1,
29               trControl=trainControl(method="cv", number=5, savePredictions=T))
30 print(model) #note that this is a quite poor prediction model
31 prediction <- predict(studyArea, model)
32 plot(varImp(model, scale=FALSE))
```

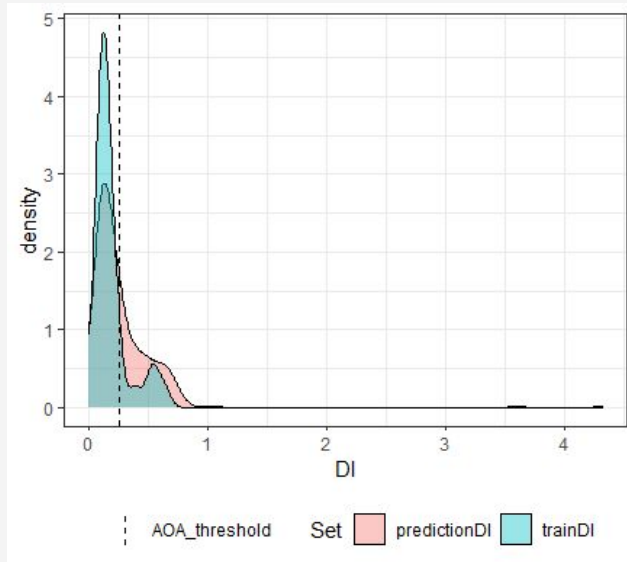
Modell trainieren - Plot



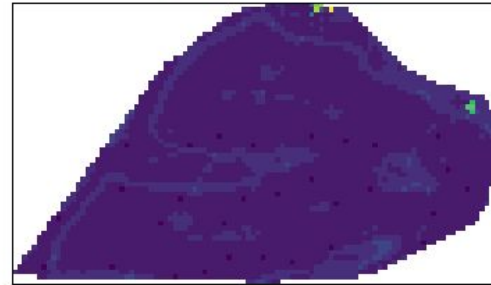
AOA des trainierten Modells berechnen

```
35 AOA <- aoa(studyArea,model)
36 plot(AOA)
37 spplot(AOA$DI, col.regions=viridis(100),main="Dissimilarity Index")
38 #plot predictions for the AOA only:
39 spplot(prediction, col.regions=viridis(100),main="prediction for the AOA")
40 spplot(AOA$AOA,col.regions=c("grey","transparent"))
```

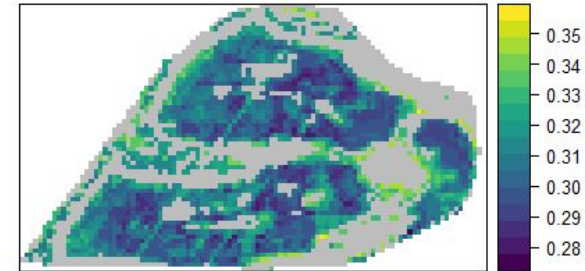
AOA des trainierten Modells berechnen - Plots



Dissimilarity Index



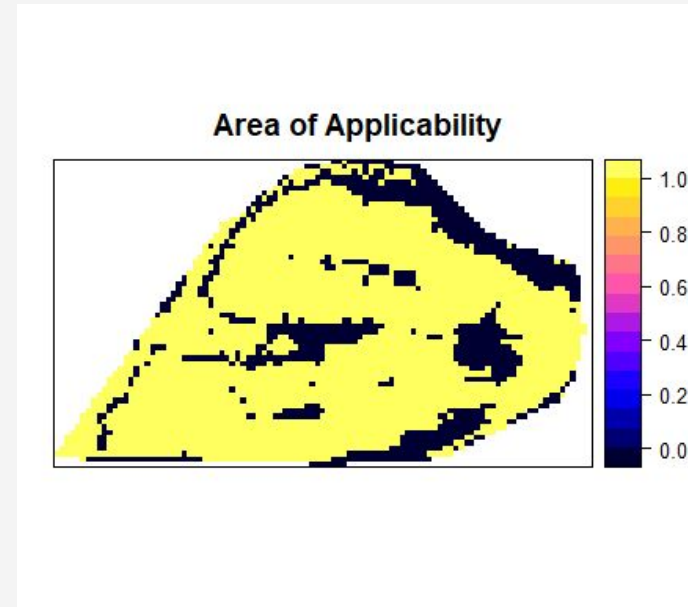
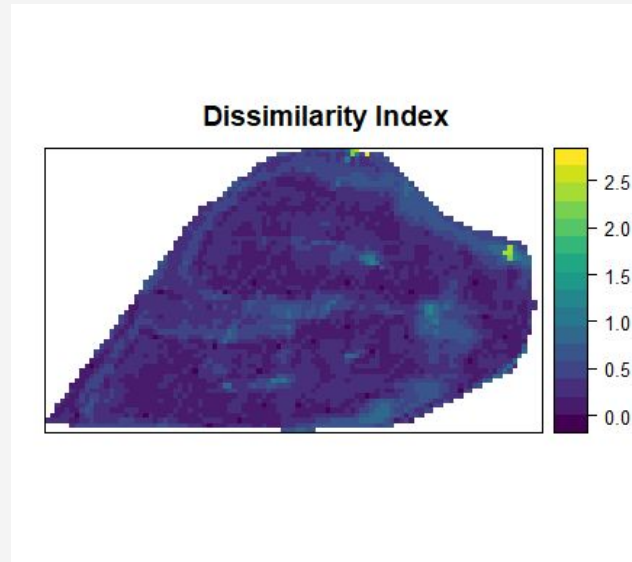
prediction for the AOA



AOA ohne trainiertes Modell

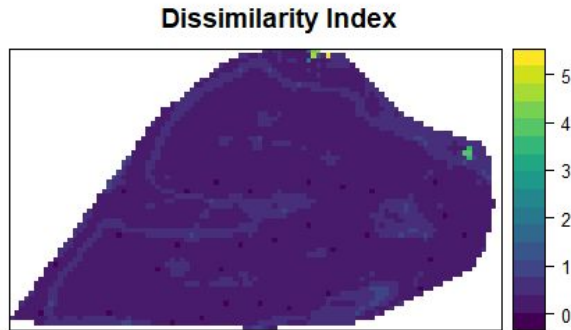
```
46 AOA <- aoa(studyArea,train=trainDat,variables=variables)
47 spplot(AOA$DI, col.regions=viridis(100),main="Dissimilarity Index")
48 spplot(AOA$AOA,main="Area of Applicability")
```

AOA ohne trainiertes Modell - Plots

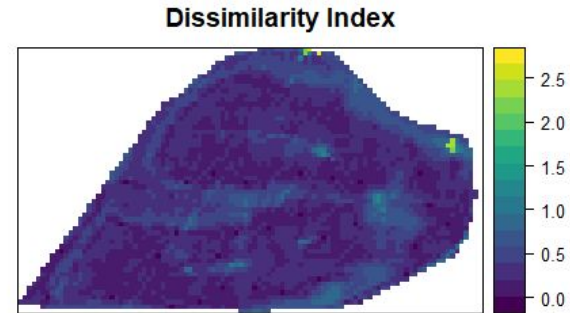


Vergleich

Trainiertes Modell



Nicht trainiertes Modell



Zusammenfassung

- AOA wird verwendet um Bereiche eines Spatial-Datensatzes zu identifizieren, in denen die Schätzungen eines trainierten Modells nicht sicher angenommen werden können
 - **AOA ist der Bereich, in welchem ein Modell mit einem bestimmten Fehlerprozent Schätzungen machen kann**
 - **Schätzungen außerhalb des AOA sollten mit Vorsicht behandelt werden**
-