Impact of multidimensional parameters on the area of applicability (AOA) for machine learning based classifications

Developing a standardized workflow of RGB data processing to test for best resulting AOA

Andreas Schönberg B.Sc.[1]

[1] Master student at the department of Geography, Philipps University of Marburg, Germany

Abstract

Machine learning algorithms have become commonly used in spatial predictions. A major problem with the results of those approaches is the estimation of the performance outside the training areas. Cross validated model accuracy values could easily lead to overoptimistic results. While target-orientated approaches can help to gain more realistic performance values one major problem still is the estimation of the applicability for unknown space. The Area of applicability (AOA) approach developed by MEYER & PEBESMA (2020) can be used to test the applicability of a model for an area of interest. It is based on the relationship of the predictor values in each cell compared to those used in the training. Therefore the AOA of a model is highly depending on the used data. I will test the impact of several different dimensions on the AOA using artificially layer (spectral indices and filter) as well as common used methods for dimensional reduction (like PCA). The aim of this study is to develop a standardizes workflow to test which data processing approach leads to best performing prediction and AOA for a given Area of interest (AOI).

Keywords: Area of applicability (AOA), machine learning, random forest, remote sensing, spatial mapping, spatial filter, spectral indices.

1. Introduction

Spatial predictions based on machine learning approaches recently have become a buzz word in remote sensing science (FOX ET AL 2017; MEYER ET AL 2016; CUTLER ET AL 2007) . In spatial classifications for example models can be trained from small training areas or even single pixel to predict the classes for much greater areas of interest (AOI). Machine learning algorithms tend to sensibility to over-fitting (MEYER ET AL 2016) due to over-sized data sets, high correlating data or layer which contains continuous data (like coordinates).

Therefore a very common used algorithm is Random Forest (BREIMAN 2001) which is known for relatively high stability and lesser risk of over-fitting. Random forest is a non-parametric ensemble approach and suffers a direct quantification of prediction error (COULSTON ET AL 2016).

For validation strategy the k-fold cross validation (CV) is a popular way for estimating the performance of a model (KUHN & JOHNSON 2013) but could lead to overoptimistic results in spatial predictions (MEYER ET AL 2016). The Leave location out cross validation (LLOCV) approach (MEYER ET AL 2016) helps to improve the estimation of model performance by leaving out data based on locations instead of random pixels. See MEYER ET AL (2016) for more detailed description for target-orientated validation strategies. Even with high performance models using the LLOCV the prediction could be useless due to other relations of the data outside the training areas. The area of applicability approach developed by MEYER & PEBESMA (2020) provides a way to estimate the spatial applicability of the prediction for unkown space by comparing the euclidian distances of the predictor in each cell with those used for the training. For a detailed description of the AOA approach see MEYER & PEBESMA (2020). The AOA therefore highly depends on the data and the relation of its dimensions. For datasets especially 3-band RGB data the processing can increase the performance of the model by giving the algorithm more value information (HUNT ET AL 2005; MEYER ET AL 2019). There are several strategies to generate Raster stacks containing additional information like the computation of artificially layers like filter and indices (HUNT ET AL 2005; MEYER ET AL 2019; FIERENS & ROSIN 1994) and principal component analysis (PCA) (ESTORNELL ET AL 2013).

Further the design of the training area will have an impact on the results (ZHAO ET AL 2020; JIN ET AL 2014) because it directly influences the amount of data values which will be used by the algorithm. Furthermore strategies like forward feature selection (FFS) (MEYER ET AL 2019) can be used to determine the best fitting layers from those Raster Stacks for the best fitting model. Those processing approaches contain several to endless variables and create multi dimensional hyperspaces where any single variable can have an impact on the results. Hereby most dimensions depend on each other or influence other processing steps. It is necessary to test the impacts of those parameters on both the model and the AOA to decide which data would lead to the best results for the AOI.

In this study I will test the impacts of several parameters from the multidimensional hyperspace on the AOA and the model performance for high spatial resolution RGB images. To compute the data sets I will use different strategies to create Raster Stacks containing artificially layers and will use different strategies to reduce its amount to handle correlations. Further I will test the impact of the training design as well as the influence of using feature selection methods (FFS). The results are highly depending on the data so they probably could not be used for any other AOI. Therefore the main goal of this study is to identify parameters which have an impact and develop a standardized workflow to easily compute several data sets and test them for the best resulting AOA and model for any given AOI.

From this I hypothesis that I am able to develop a workflow to test the impact of several selected dimensions for the best resulting classification and the respective AOA. Further I hypothesis that I can reveal the relationship between the dimensions and distinguish which have a significant impact. I assume that the workflow is capable of computing several data sets in a comprehensible and reproducible way and further provide an easy way to test the data sets in a standardized setup.

At least I will deliver a bundle of ‘R’-functions to enable the usage of this workflow for any given AOI.

2. Data and Methods

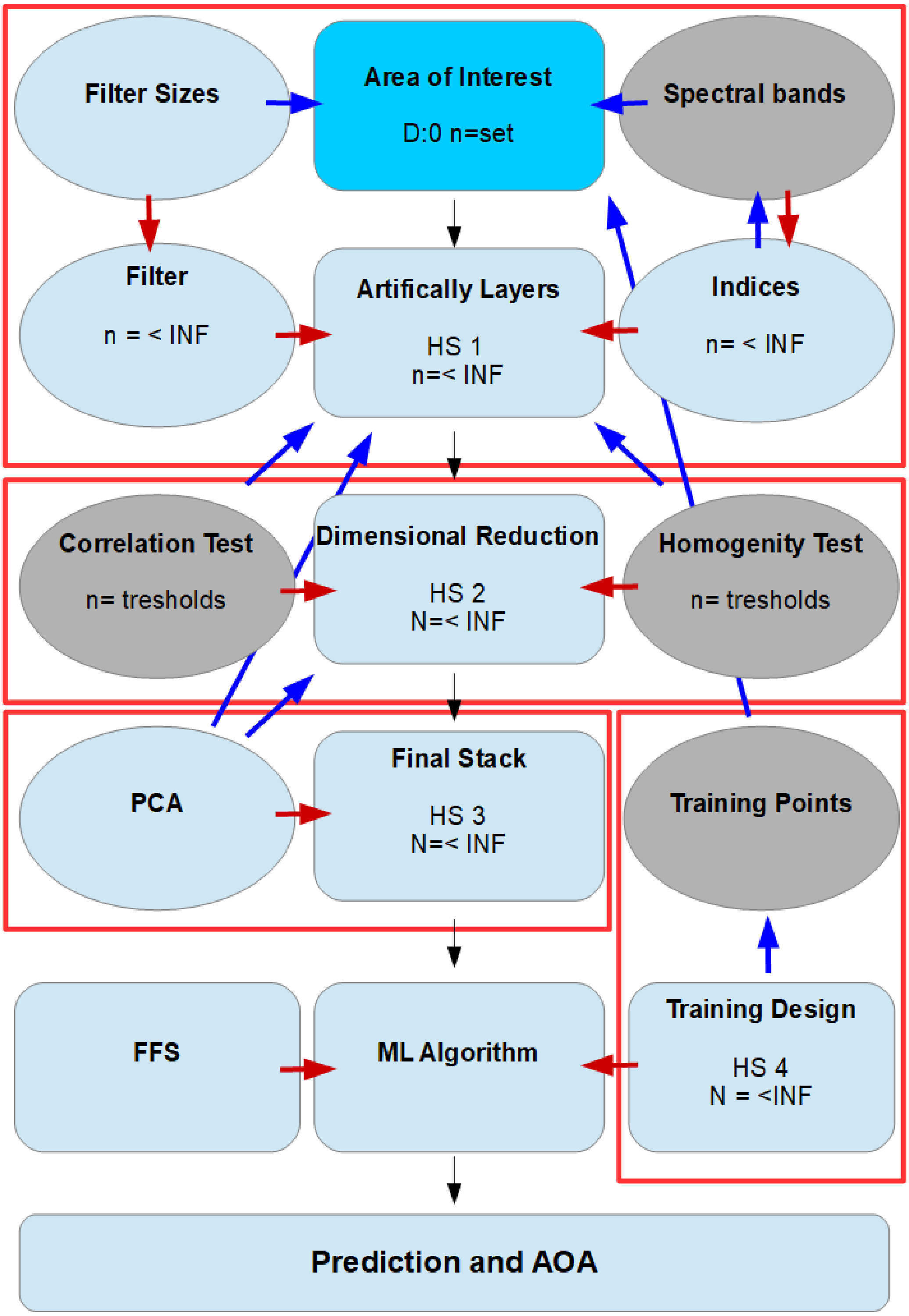

Fig.1 : *Schematic "map" of the multidimensional hyperspace: Each Dimension has multiple to infinite variables and depends on others (blue arrow) and increases or multiplies the amount of resulting variables in the hyperspace (red arrow). Each processing step (black arrow) merges individual hyperspaces and increases the resulting variables. Grey dimension are set to fixed variables to decrease the total amount.*

There are nearly infinite variables which could have an influence on the performance of a model and the AOA. Most variables depend on each other and create several multi dimensional hyperspaces (Fig. 1). Each dimension multiplies or increases the amount of total variables. While some dimension have natural (mathematic or logical) ranges other could be extended endless tending to infinite variables. For this study I will reduce the amount of those dimensions by selecting common used variables as well as setting some dimension to only one variable. Further I will use costume build R-Packages to test several ways for the data processing and deliver a comprehensible and reproducible workflow to test models for a best resulting prediction with high AOA. The data processing and modelling is performed in R version 4.0.4 (R Core Team 2020). The used Image and R-Scripts can be retrieved from https://github.com/SchoenbergA/AOA-Impact.

2.1 Study Area and Training Design

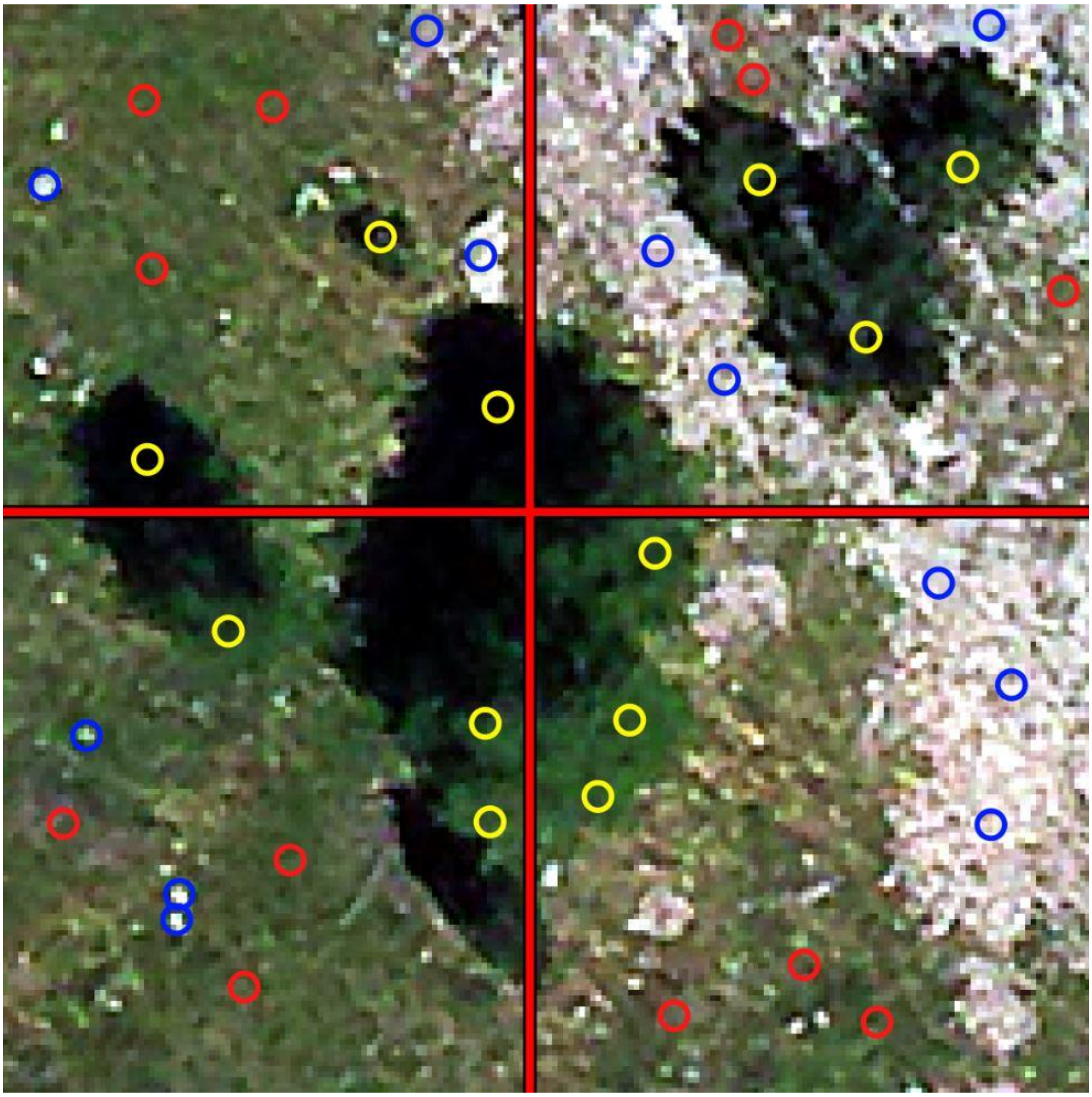

Fig. 2 : *The Study Area with position of training points and border for the sctros used for the LLOCV. The training points (yellow= tree, blue= soil, red=grass) are located in the center of the cycles which are used for overview in this image only.*

The first set of variables is the Area of Interest (AOI) and the training design. For the development of the RGB-workflow I will use a small aerial image with 0.15 meter spatial resolution (see fig.2). The scene shows some trees surrounded by grass and open soil located in the French alps representing the three classes: trees (t), grass (g) and soil (s). While soil is separated by its gray to silver colors from the other classes the trees and grass share dark to brighter greens. Due to the spectral similarity of tree and grass the scene is perfectly to test the ability of the algorithm to detect the borders depending on the datasets. For a more standardized testing and to reduce subjective decisions by greater polygons I will use points for the training. I would not test other sets of training areas but modify its ranges and forms. To implement the LLOCV approach I will cut the scene into four sectors (NE,SE,SW,NW) and set three training points for each class in each sector. The points are set on positions representing the average behavior of the classes. Both the position of the training areas and the design in form and size have essential influence on the data which will be used for the classification. Combining those variables creates a hyperspace (fig X HS4) with different positions of the training areas can be combined with different forms as well as sizes of the areas. Due to the high spatial resolution a class can be represented by a wide range of values which may not be captured by single cell training points. To test the influence of those parameters I will use spatial buffers around the points with same ranges for each point. Beside cycles (c) I will although test the influence of rectangles (r).

I assume smallest areas would contain to less information while too wide areas would catch values of other classes. Especially small spots of soil surrounded by grass could catch too many values representing grass as well as training points in close neighborhood would overlap and catch values multiple times on which the ML algorithm probably would react.

2.2 Artificially Layer Computation

The next hyperspace of variables (fig.1) is based on the computation of the artificially layers (ALs) where different types (filter, indices) multiplying with different sizes which depend on the image resolution lead to nearly endless variables. Besides the original red, green and blue bands I will use ALs to compute additional information. There are nearly endless options to compute artificially layers. Too many layers would take long processing time and could lead to over fitting. So it is necessary to limit its amount and focus on those ALs which could probably improve the performance. On the other hand in this study I want to reduce subjective decisions to receive most neutral results. I assume for high resolution RGB images filter and indices are most recommended. Due to the high resolution the cell values could be highly different even in close neighborhood. So I assume that filter would probably increase the performance of a classification by smoothing (FIERENS & ROSIN 1994) the values in close neighborhood of the training areas as well as for the prediction. The filter requires a Moving Window defining the amount of cells in neighborhood which the filter function is assigned to. Indices are common used in remote sensing approaches (RAY ET AL 2004 ;HUNT ET AL 2005; MEYER ET AL 2019) to detect vegetation and could help to detect and separate non-vegetation (soil) (RAY ET AL 2004) as well as the vegetation classes (grass/trees) due the sensitivity to chlorophyll (HUNT ET AL 2005; HUNT ET AL 2013) .

For the computation of the artificially layers I will use a custom bundle of R-function 'LEGION' (https://github.com/SchoenbergA/AOA-Impact) which provides general spatial filter based on the 'Raster::focal' function (HIJMANS 2020) and sobel filter for edge detection (Table 1) as well as common used indices (Table 2). For further information about the spectral RGB indices see: THE IDB PROJECT (2020); RAY ET AL (2004); HUNT ET AL (2013).

Table 1: *Filter functions provided by 'LEGION'.*

Name	Calculation
Sum	sum of all cells in a MovingWindow
Minimum	minimum value of all cells in a MovingWindow
Maximum	maximum value of all cells in a MovingWindow
Mean	mean value of all cells in a MovingWindow
Standard deviation	standard deviation of all cells in a MovingWindow
Modal	most frequent value of all cells in a MovingWindow
Sobel	sobel edge detection filter in horizontal and vertical directions
Sobal horizontal only	sobel edge detection filter in horizontal direction only
Sobal vertical only	sobel edge detection filter in vertical direction only

Table 2: *Indices implemented in 'LEGION' and its calculation.*

Name	Tag	Calculation
Visible Vegetation Index	VVI	(1 - abs((red - 30) / (red + 30))) * (1 - abs((green - 50) / (green + 50))) *(1 - abs((blue - 1) / (blue + 1)))
Visible Atmospherically Resistant Index	VARI	(green-red)/(green+red-blue)
Normalized Difference Turbidity Index	NDTI	(red-green)/(red+green)
Redness index	RI	(red**2/(blue*green**3)
Soil Colour Index	CI	(red-green)/(red+green)
Brightness Index	BI	sqrt((red**2+green**2+blue*2)/3)
Spectra Slope Saturation Index	SI	(red-blue)/(red+blue)
Primary Colours Hue Index	HI	(2*red-green-blue)/(green-blue)
Triangular Greenness Index	TGI	(-0.5*(190*(red - green)- 120*(red - blue))
Green Leaf Index	GLI	(2*green-red-blue)/(2*green+red+blue)
Normalized Green Red Difference Index	NGRDI	(green-red)/(green+red)

Using this indices and filter functions with theoretically endless moving window sizes would end up in an artificially computed technically endless multidimensional hyperspace with endless combinations of possible Raster Stacks to compute. Further the processing time could be very long. Using all of this 9 filter functions on the 3 original bands as well as on all 11 indices would result in 9*3+9*11 = 126 artificially layers for every moving window size. Depending on the extent of the AOI the resolution of the image and the environmental behavior much sizes could be value to test. Even if only using sizes of 3, 5, 7 and 9 would result in 126*4 = 504 artificially layers.

To test the influence of the used variables I will compute several Raster Stacks (Table 3) for the test of the impact on prediction and AOA and further test which processing workflow for the data sets would lead to the best results. First I will use only the red, green and blue bands tagged 'rgb' without any AIs as a base to compare to. Next I will compute all indices and all filter for the indices and the single RGB bands using moving windows of 3, 5, 7 and 9 respectively named 'full'. Moving windows of greater than 9 (9*0,15m = 1,35m) seems problematic with the test area especially for the soil training points in the south. Further I will compute one with smaller moving window of 3 and 5 ('small') which would probably be more precise.

To test the influence of subjective pre selection I will at least compute a Raster Stack ('selected') by dropping some indices and filter. Assuming the RGB spectral indices are highly correlated at all and further the RI and HI are lesser useful on the target area I select the indices: VVI, NDTI, GLI and NGRDI. For the filter function the MAX and MIN probably lead to homogeneous clusters especially with wider moving window sizes. Therefore I will use all filter functions except MIN and MAX.

Table 3: *Raster Stacks with computed artificially layers.*

Raster Stack	Computation workflow
rgb	The original red, green and blue bands
full	All indices + rgb with all filter 3,5,7,9 + rgb and indices (original)
small	All indices +rgb with all filter 3 and 5 + rgb and indices (original)
selected	Indices: VVI, NDTI , GLI , NGRDI ;Filter sum, mean, modal, sobel, sobel_hrzt, sobel_vert + rgb and indices (original)

2.3 Dimensional Reduction

There would be a high amount of correlating layers in the Raster Stacks due to the mechanism of computation which could lead to over-fitting when used for the machine learning. It is recommended reducing this hyperspace (Fig. 1) in further processing to only those layers providing value information.

To reduce the dimensions in the Raster Stacks I will test for correlation and delete highly correlating layers. Further some layers especially the filtered indices could result in highly homogeneous layers contain only a few values. Those layers could lead to explaining the dependencies of the training area very easy for the algorithm and lead to false predictions. To handle those layers I will further test the distribution of the data value and drop layers by thresholds. Both methods itself create another multidimensional hyperspace (HS2) with nearly endless possible threshold values for correlation and homogeneity leading to different output Raster Stacks as well as the order in which both methods are used. To reduce the amount of dimensions I will use fixed thresholds for both approaches based on experiences. For the correlation test I will drop all layer with a cor value of >= 0.7 and <= -0.7 using a threshold of 0.7 for the function 'LEGION::detectRstCor' while thresholds of 0.9 for a value range of 0.1 will drop all layers which have >= 90% of data values in <=10% of the data range using 'LEGION::detectRstHmgy'. To test the impact of the order for using the methods I will first use the correlation test on each Raster Stack (Table 4) and then the Homogeneity test (adding prename 'hmgy_' for last method used) as well as the other way around by first using the Homogeneity test and then dropping remaining layers by using the correlation test (adding prename 'corT_' for last method used)

Table 4: *Raster Stacks with dimensional reduction.*

Raster Stack	AIs computation	Dimension reduction
hmgy_full	All indices + rgb with all filter 3,5,7,9 + rgb and indices (original)	First correlation test, than test remaining layers for homogeneity
hmgy_small	All indices +rgb with all filter 3,5 + rgb and indices (original)	First correlation test, than test remaining layers for homogeneity
hmgy_selected	Indices: VVI, NDTI , GLI , NGRDI Filter sum, mean, modal, sobel, sobel_hrzt, sobel_vert + rgb and indices (original)	First correlation test, than test remaining layers for homogeneity
corT_full	All indices +rgb with all filter 3,5,7,9 + rgb and indices (original)	First test for homogeneity, than test remaining layers for correlation
corT_small	All indices +rgb with all filter 3,5 + rgb and indices (original)	First test for homogeneity, than test remaining layers for correlation
corT_selected	Indices: VVI, NDTI , GLI , NGRDI Filter sum, mean, modal, sobel, sobel_hrzt, sobel_vert + rgb and indices (original)	First test for homogeneity, than test remaining layers for correlation

2.4 Principal Component Analysis (PCA)

Another strategy to generate information for the ML algorithm is the principal component analysis (PCA) which is commonly used for data processing (ESTORNELL ET AL 2013; DESMER ET AL 2013; CAO ET AL 2003). It is used for dimensional reduction by returning the values which are orientated near a line within the euclidean space. The PCA methods leads to again several variables like the combination with other layers and the amount of PCAs to use (HS4) The first PCA (PCA1) is orientated at a line with the mean of the multidimensional space of the data and contains the most information from every input layer. One limitation of this technique is that with the resulting PCA layer the user is not able to conclude the influence of the single input layers. I will use the 'rasterPCA' function provided by the 'Rstoolbox' package (LEUTNER AT AL 2019) and use the first PCA only (PCA1) for the Raster Stacks (Table 5). First I will use the PCA for the original rgb bands ('pca_rgb') and further I will compute a PCA for the rgb bands and the indices ('pca_ind') as well as for only the indices ('pca_ind_only') assuming that the information from the RGB bands are already containing in the indices. The computation of HI and RI can lead to infinite values which would not work in a PCA (LEUTNER AT AL 2019). Therefore I skip these two indices for the PCA approach. Furthermore I will use the RGB bands and all filter for the RGB bands with moving window sizes of 3, 5, 7 and 9 to compute a PCA ('pca_rgb_filt'). At least for comparison purposes between the dimensional reduction techniques of PCA and the correlation/homogeneity approach I will calculate PCAs for each of the AIs Raster Stacks.

Table 5: *Raster Stacks with principal component analysis.*

Output Raster name	**Input Raster Stack for PCA**
pca_rgb	RGB bands
pca_ind	RGB bands + all indices (without RI and HI) for RGB
pca_ind_only	All indices (without RI and HI) for RGB
pca_rgb_filt	All filter with 3, 5, 7 and 9 for RGB and added RGB bands
pca_ hmgy_full	hmgy_full
pca_ hmgy_small	hmgy_small
pca_ hmgy_selected	hmgy_selected
pca_ corT_full	corT_full

2.5 Forward Feature Selection (FFS)

The dimensional reduction always leads to a loss of information by dropping parts of the data. Instead of leaving out parts of the data an automated selection would have access to all the data. For the machine learning approach the algorithm would use all layers in the input Raster Stack to compute a single model. A forward feature selection (FFS) like implemented for ‘R’ by MEYER (2018; 2020) would compute models with each layer in the Raster Stack and select the best fitting layers. This technique could help to avoid information loss and increase the performance of a model. I will use the 'CAST::ffs' function from the 'CAST' package (MEYER 2020) where models are trained first for every possible pair of two layers and kept the best model. Based on this model the layers are iteratively increased to test for an improvement in model performance and stops if none of the remaining layers would further increase the current best model (MEYER 2020).

With the selection of layers the resulting models and predictions probably differ from those based on the unselected fully used Raster Stacks. Therefore I will test some Raster Stacks with the FFS to compare the results with the other methods and detect differences to estimate the best workflow.

First to compare the layer selection with the results for the PCA I will compute Raster Stacks (Table 6) with the same layers which have been used to calculate the respective PCA: 'ffs_ind' and 'ffs_ind_only'. Using a FFS on only the RGB bands seems quiet useless and will not be tested. Additionally I will test if the skipped indices RI and HI would be selected by a FFS by computing a Raster Stack with all indices 'ffs_ind_all'. To further compare the FFS with the PCA including the filter layer the Raster Stack would result in 108 layers which would take a long time for the FFS to process. To reduce the time for this study I will use the correlation approach to reduce the amount of layers ('ffs_filtcor'). At least I will use the FFS to test the impact of a PCA layer in a Raster Stack by calculating a PCA for the 'ffs_filtcor' Raster Stack and add the frist three PCA to the Raster Stack ('ffs_pca'). Probably the PCA is selected and changes the whole selection.

Table 6: *Raster Stacks with artificially layers for forward feature selection method.*

Raster Stack	Input Raster Stack	Equal layer used for PCA
ffs_ind	Rgb bands + all indices (without RI and HI)	pca_ind
ffs_ind_only	All indices (without RI and HI)	pca_ind_only
ffs_ind_all	All indices (including RI and HI)	-
ffs_filtcor	All filter for rgb bands, reduced by correlation	pca_rgb_filt (*)
ffs_pca	ffs_filtcor + PCA for ffs_filtcor	-

2.6 Machine Learning Approach

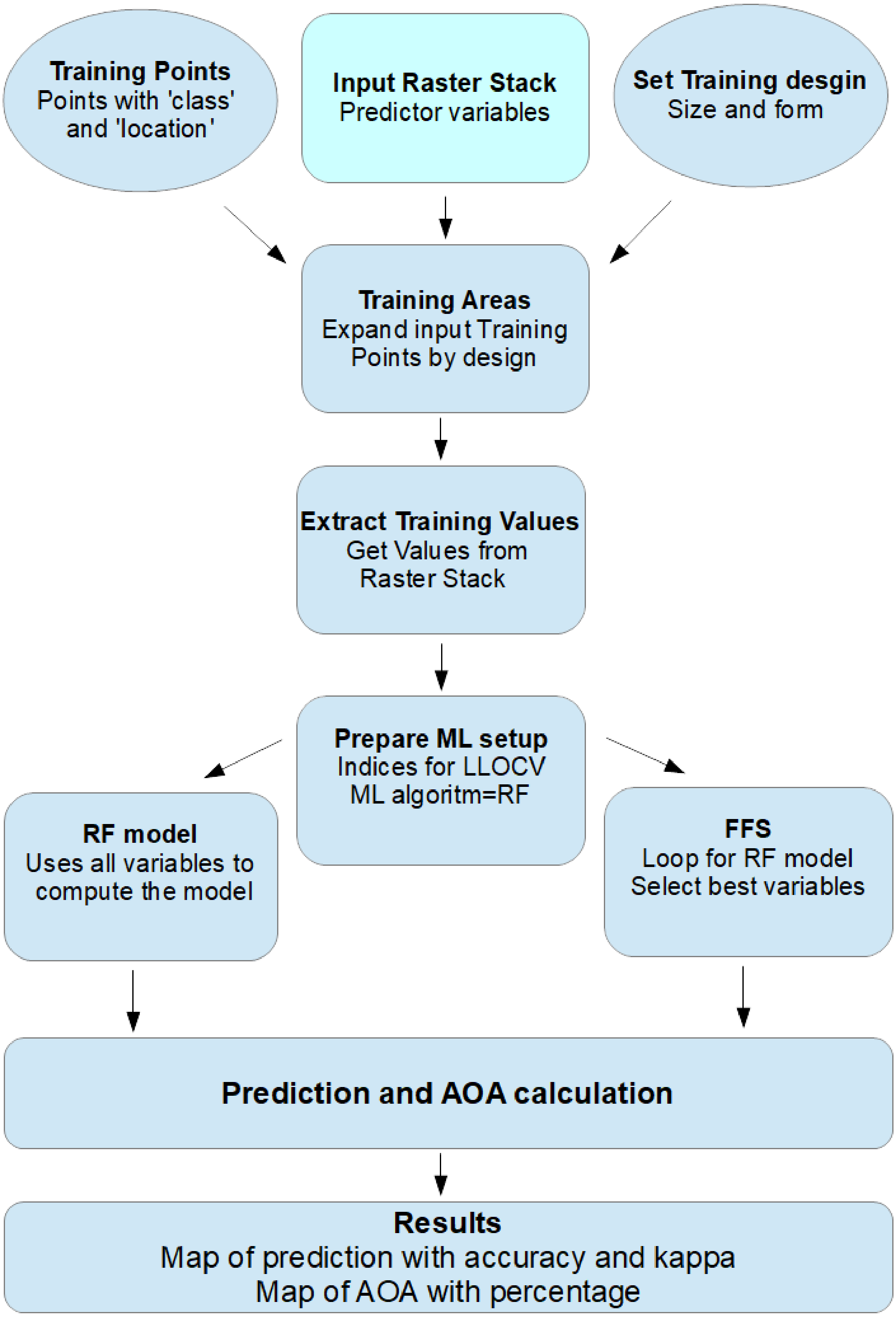

Fig. 3: Schematic method for 'IKARUS_dawn' wrapper function.

The processed Raster Stacks will now be used to compute models and prediction as well as to calculate the respective AOA. Therefore I will use my costume build wrapper function 'IKARUS_dawn' based on my function bundle 'IKARUS' (https://github.com/SchoenbergA/AOA-Impact) which is mainly a wrapper for 'CAST' (MEYER 2020) with additional strategies for the extraction of training data. The 'IKARUS_dawn' function is designed to perform all necessary steps in one. First the training data for the respective training design is extracted by 'IKARUS::exrct_Tdat' from the input raster Stack. Therefore cycle or rectangle polygons are created with a selected radius (in meter) from the position of the training points. The data of all layers within the input Raster Stack is extracted and will be used for further steps. Next the LLOCV is prepared using 'CAST::CreateSpacetimeFolds' (MEYER 2020) by defining Indices for the data to be left out based on the four sectors of the training design. Depending on the method of using all layers or a FFS the random forest algorithm is started and the resulting model will be used to predict the final classification map. Further the AOA is calculated using the function 'CAST::aoa' (MEYER 2020). At least the function will print the classification and a map for the AOA and additional print the resulting accuracy and kappa for the model and the percentage of the AOA for comparison. If a FFS is used the selected variables are printed respectively.

The resulting values of accuracy and kappa along with the AOA percentage for the area will explain the performance of the prediction depending on the used training data. So even an accuracy of over 90 percent must not mean a nearly perfect prediction at all. Therefore an automated evaluation of the results using the performance values for the classification seems lesser recommended than a visual evaluation. But some dependencies of the dimensions should be explained by the relationship of those values. Due to the high resolution and fine distribution of class ‘grass’ and ‘stone’ in close neighborhood I did not use response data for evaluation. Likewise the results for the AOA in percent require a visual evaluation to estimate the spatial distribution because even lesser AOA results could occur for better prediction. For example a concentration of pixels out of the AOA could be more usable in total than a prediction with higher AOA values overall but wider spread of single pixels which are not usable.

2.6 Test-Series

For the test series the computed Raster Stacks will be used to compute the classification and the respective AOA (Table 7). While the amount of total dimension has been reduced by the selection of some variables or the use of fixed parameters for others there is still one major hyperspace created by the combination of the Raster Stacks with different training designs. Testing each Raster Stack with a series of sizes and forms would end up in a huge amount of tests (n Raster Stacks*n sizes* 2 forms).Therefore I will first test the general influence of the training design with the RGB bands only and then use a single design for each Raster Stack to compare its impact based on the data only.

For the sizes in the training design test series I will use multiples of the raster cell size of 0.15 meter: 0.15, 0.3, 0.6, 1.2 for both rectangles (r) and cycles (c) forms. Wider radius seems unnecessary to test because here the training areas would be tend to catch two or more classes and further would lead to a huge amount of data which would slow down the algorithm.

For the test series using each computed Raster Stack I will use a fixed training design of cycles in 0.3 meter radius (c03). Finally I will use the training designs of cycles with 0.15, 0.3, 0.6, and 1.2 radiuses for the visually best performing Raster Stack for fine tuning.

Table 7: *Test series and used Raster Stacks with respective training designs.*

Test series	predictors	Training design
Impact of training design (general)	RGB bands	Cycles and rectangle in sizes of 0.15, 0.3, 0.6, 1.2
Impact of AIs and dimensional reduction methods	Hmgy and corT Raster Stacks	Cycles in 0.3
Impact of PCA	PCA Raster Stacks	Cycles in 0.3
Impact of FFS and comparison of PCA dimensional reduction with FFS	FFS Raster Stacks	Cycles in 0.3
Fine-tuning	Visually best performing Raster Stack	Cycles in sizes of 0.15, 0.3, 0.6, 1.2

3 Results

In total I have tested 28 combinations of Raster Stack and training design. For the prediction the accuracy and kappa of the model is printed as well as the percentage of the AOA.

3.1 Impact of Training Design

There is only a slightly difference between the cycles (Fig. 4) and rectangle (Fig. 5) designs despite for which sizes. Overall the rectangles have minimal lesser accuracy compared to the cycle forms except for the 0.15 meter sizes where the rectangles are slightly higher in performance. The performance for the rectangles decreases with increasing size of the training area while the cycles show the highest accuracy (0.9177) for the c30 design which is minimal higher than the c15 with 0.9162. For greater sizes the accuracy deceases like for the rectangles. Both designs look very equal for the tree class and differ a little for the soil and grass classes. The respective AOA is very high overall from 98.28% (r120) to 99.67% (c15). The AOA constantly decreases for both design with increased sizes of the training areas for <0.1%. The pixels which are out of the AOA are distributed over the full scene for every design and can often be located at the outer line of the trees. Independent of the high accuracy for the models the predictions look less well performing due to huge amount of pixels of a class mixed up with another. With 0.15 and 0.3 meter radius there are lots of tree pixels allocated within the grass ones while the greater sizes lead to much more stone class pixel.

3.2 Impact of Artificially Layers and Dimensional Reduction

Except the 'selected' one all artificially layer Raster Stacks contain NA values and lead to empty cells for both the prediction and the AOA (Fig. 6). While the 'Cort_full' and 'hmgy_small' lead to higher accuracy compared to only using the RGB bands the visual evaluation shows huge rectangle cluster of classes and for 'CorT_full' a lot of NA pixel. At least the 'hmgy_selected' seems to lead to a well performing prediction with lesser single pixel of stone and tree within the grass. Further the results for the 'hmgy_small' look quietly precise with lesser wrong classified trees and more clustered stone pixels but contain NA areas. None of the models reaches the AOA percentage of the RGB only result of 99,28% but the distribution is much more concentrated on areas compared to the overall single pixels for RGB bands only. While the 'CorT_full' has huge areas out of the AOA located most in the areas of the tree-shadows the other results show pixel out of the AOA concentrated on the outer borders between trees and other classes in North West directions.

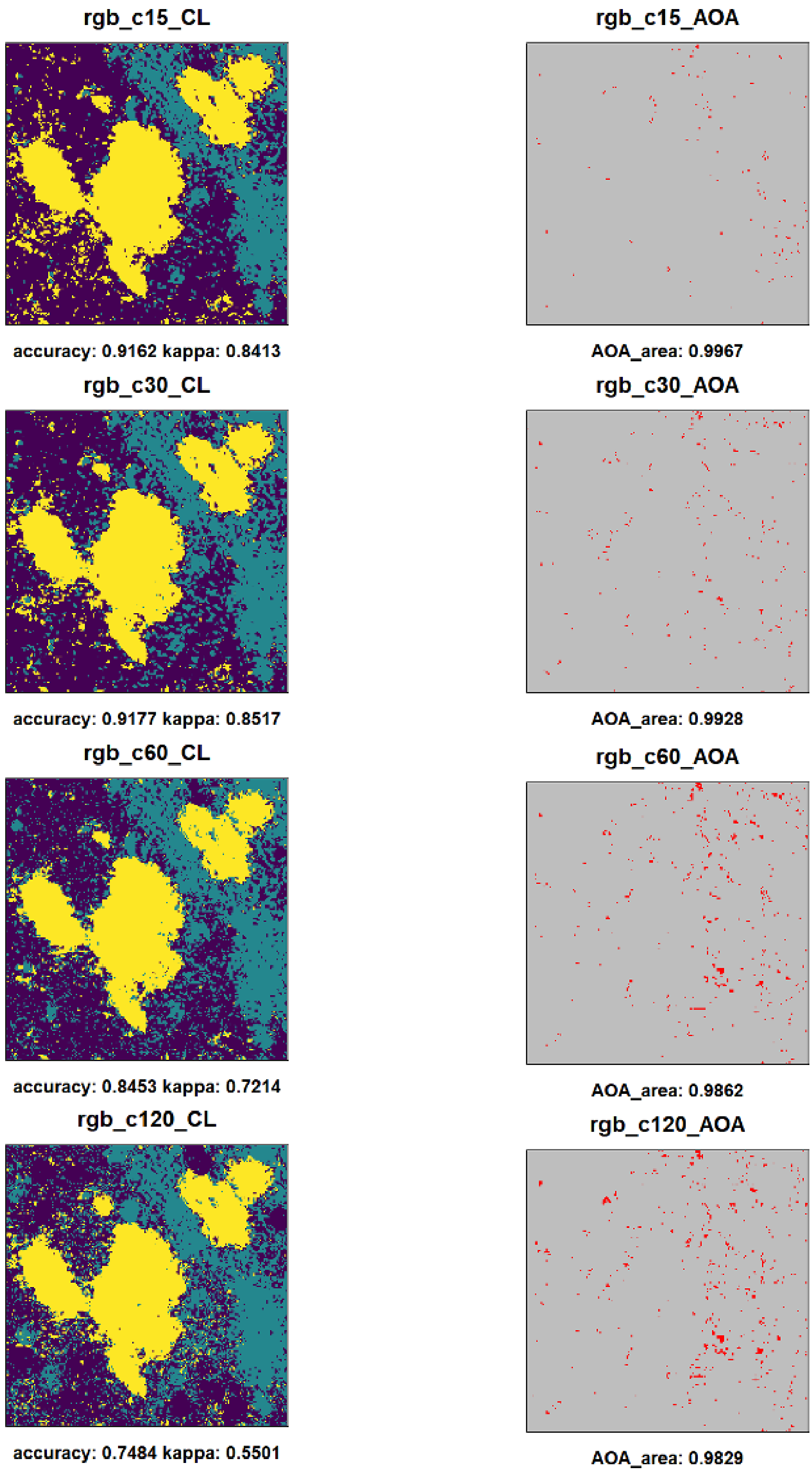

Fig. 4: *Prediction and AOA for test-series "*Impact of training design (cycles)*". Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red).*

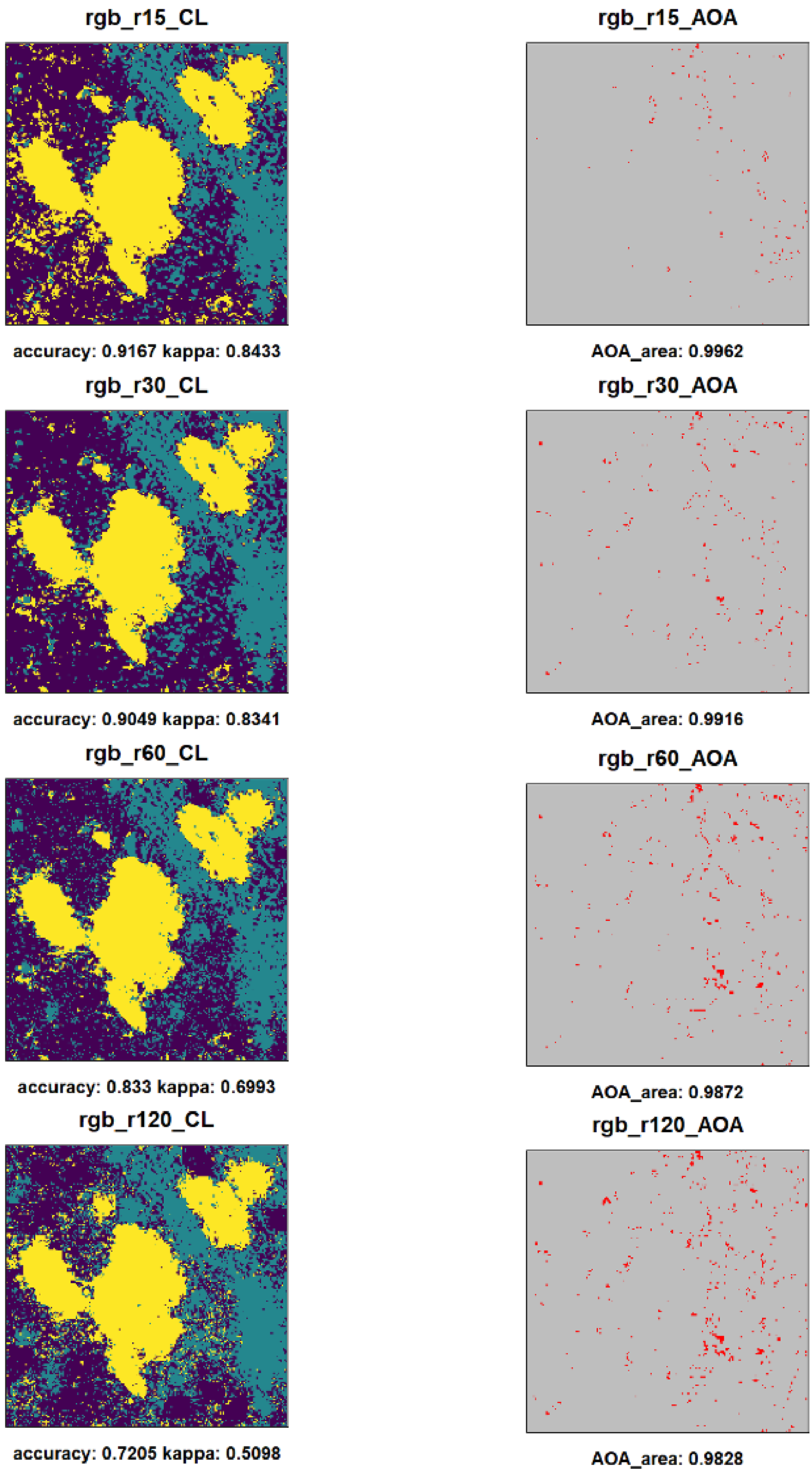

Fig. 5: *Prediction and AOA for test-series "*Impact of training design (rectangles)*". Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red).*

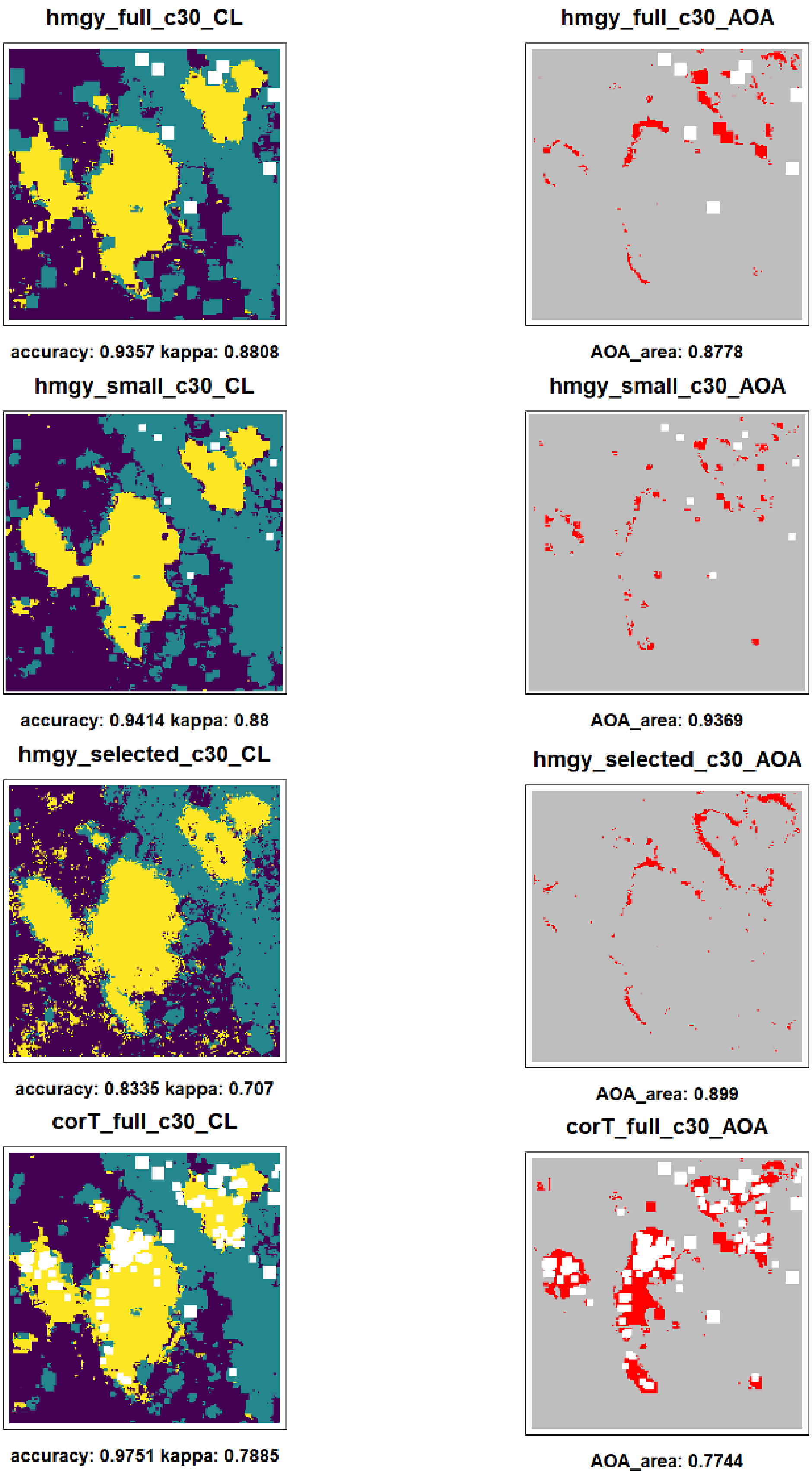

Fig. 4: *Prediction and AOA for test-series "*Impact of dimensional reduction) *Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red).*

3.3 Impact of PCA

The results for the PCAs computed from the artificially layer Raster Stacks lead to significant lesser accuracy (Fig. 7) compared to the models using the respective Raster Stack (Fig. 4) decreasing from almost around 0.90 to 0.70. Further the visual evaluation shows a far lesser separation of the classes. The results for the AOA are mostly equal in total amount but the pixels out of the AOA are spread more over the area and lesser concentrated. For 'pca_hmgy_small' and 'pca_hmgy_selected' now the outline of the tress or respectively the border between tress and no trees is most times out of the AOA. Like with the original Raster Stacks there are NA clusters except for the 'pca_hmgy_selected'. Further the variant of first using the homogeneity test before using the correlation test method results in lesser accurate results compared to the other way around.

The use of the PCA on indices without filter lead to fully different results (Fig. 8). Both the PCA using the indices with and without addition of the RGB bands as well as the PCA only using the RGB bands result in very equal predictions. All those three predictions look very equal to each other as well as very equal to the results for only using the RGB bands for the model. The accuracy for the PCA using RGB bands only is slightly higher with 0.9337 compared to 0.9177 for RGB bands only while the PCAs using the indices with and without RGB bands have slightly lesser accuracy (0.8896 for only indices and 0.9056 with additional RGB bands). The use of the PCA on the RGB bands leads to a minimal lesser AOA compare to only using the RGB bands but both PCAs with the indices result in significant lesser AOA with nearly change in the prediction performance.

The PCA computed from the filter of the RGB bands (‘pca_rgb_filt’) has a significant impact on both the prediction and the AOA. Compared to all other results the prediction seems to be the most actuate with a more clearly separation of the classes and lesser spread of single pixel spread over the area. While there are again wrong predicted tree areas in the south west sector these are accumulated instead of wide spread of single cells. On the other hand the stone areas in the north east sector seem to be greater than seen in the original image with lesser grass spots between it. Like for the prediction the AOA has a totally different distribution compared to the other results. With a relatively lesser AOA of 92.03% the cells out of the AOA are accumulated to three greater clusters and a small amount of tiny clusters located most in the north east sector. There is a significant lesser spread of areas out of the AOA with only minimal single pixels.

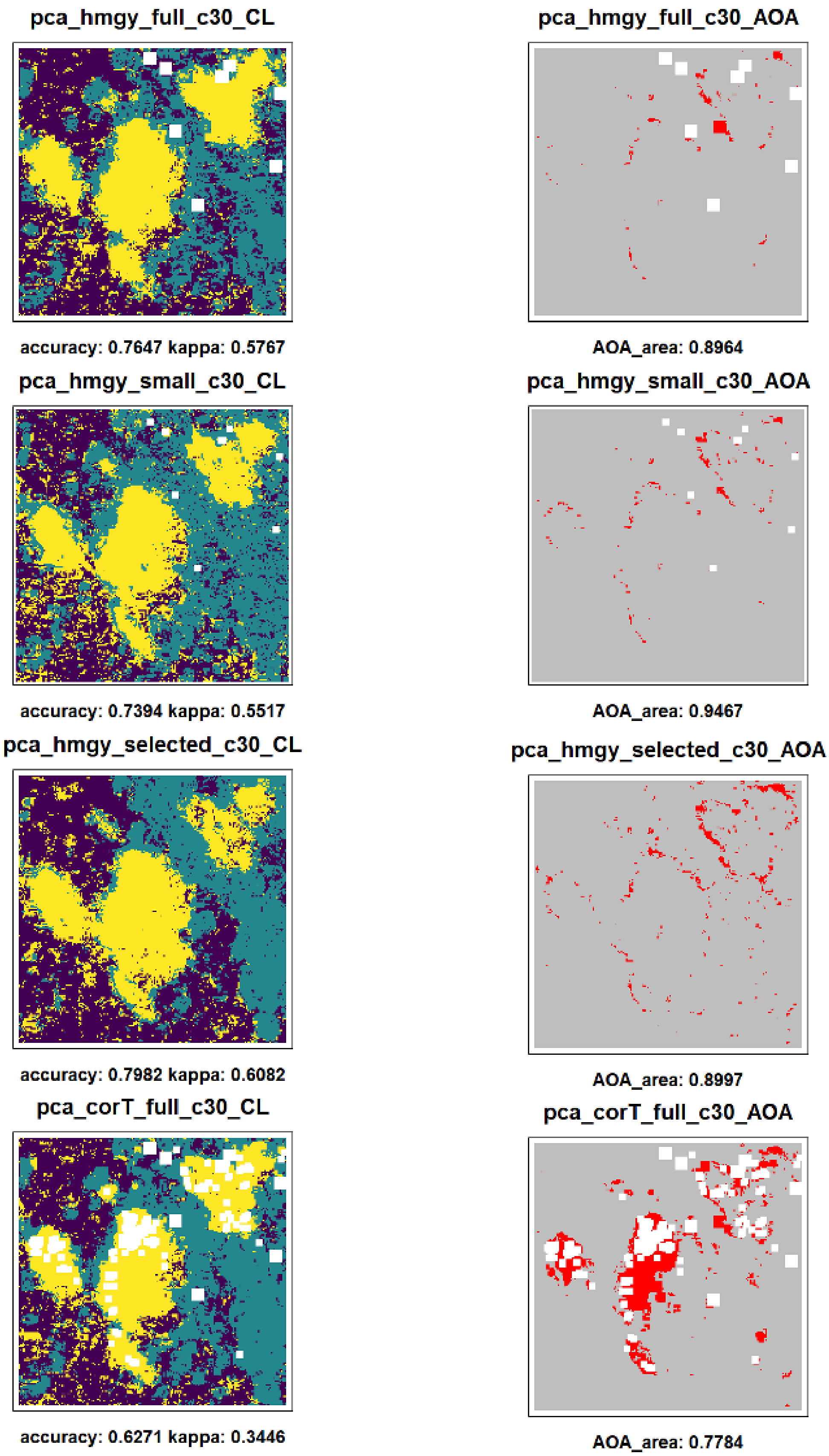

Fig. 7: *Prediction and AOA for test-series "*Impact of PCA*". Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red).*

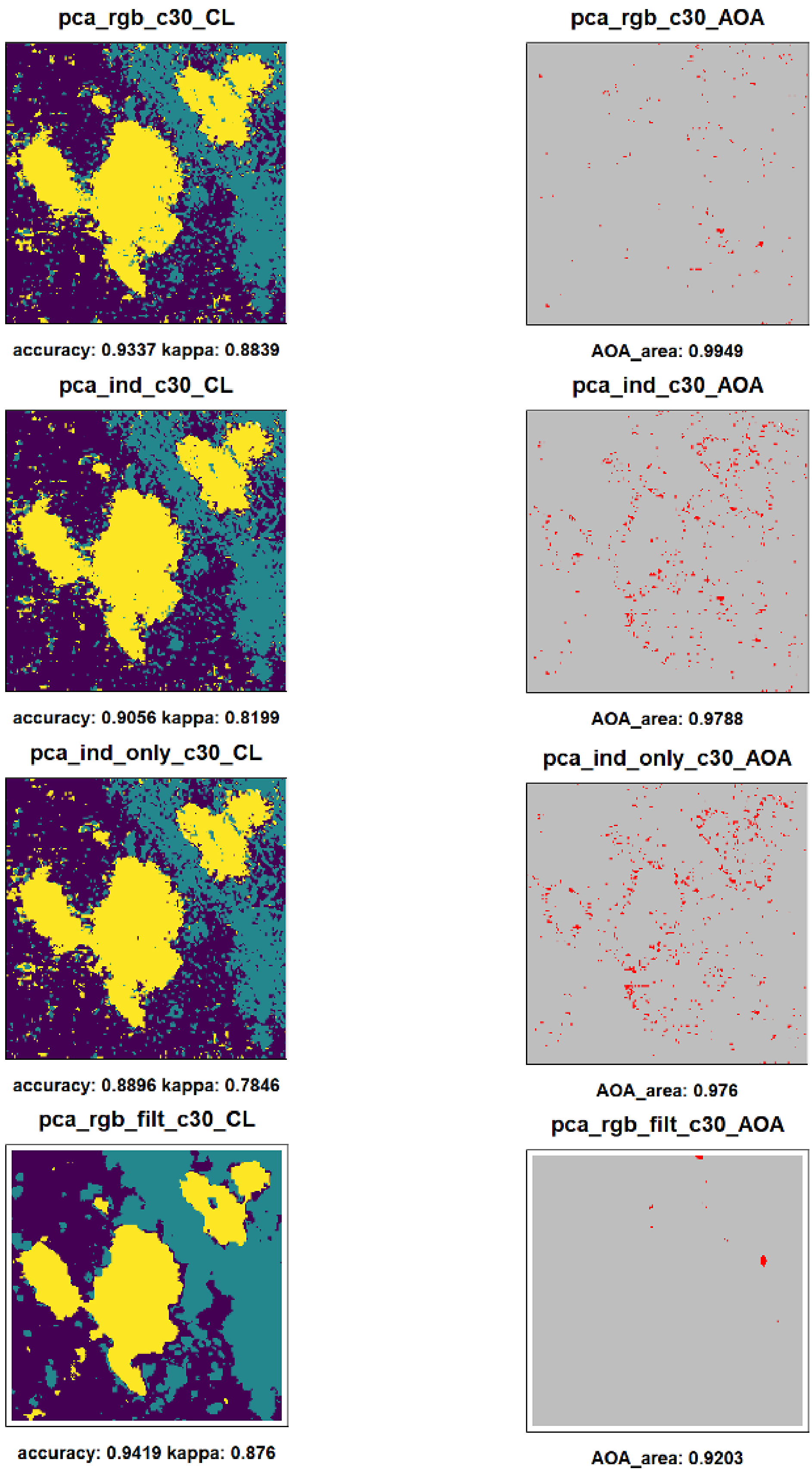

Fig. 8: *Prediction and AOA for test-series "*Impact of PCA*". Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red).*

3.4 Impact of FFS

The use of the FFS with the Raster Stacks containing the indices compared to the PCA for the indices lead to mostly equal results in visual evaluation with slightly higher accuracy overall with the FFS (see Fig. 9). The PCA for only indices compared to the FFS with the ‘ffs_ind_only’ Raster Stack lead to very equal results. Further the results for the FFS with additional RGB bands differ only minimal. For both Raster Stacks the VVI, SI and TGI indices are selected (fig XX). Furthermore the use of all indices (inclusing the HI and RI) lead to very equal visual evaluated performance but although there are NA areas due to the selected RI and HI. The result for the correlated filter Stack (‘ffs_filtcor’) lead to different results (see Fig. 10) and seems to be very accurate compared to the original image with more realistic grass spots within the stone in the north east sector. The prediction even looks more accurate than the PCA result for 'pca_rgb_filt' but has a significant wider spread of pixels out of the AOA and a lesser AOA overall (89.52 compared to 92.03). The pixels out of the AOA are distributed mostly around the trees and at the border between trees and other class and or the tree-shadow. At least if the PCAs 1 to 3 are added to the 'ffs_filtcor' the selection is completely different except for ‘red_model9’. The second PCA is selected along with "blue_min5" and "blue_modal5" instead of "green_sobel9" and "green" (see Table 8). The resulting prediction reaches high accuracy for the model and has overall wider areas for the trees compared to the 'pca_rgb_filt' result. Further there is again more grass within the stone in the north east which seems to be more realistic. On the other hand the small stone areas in the south looks like clusters with rectangle form and further there is a border of grass around the trees in the middle of the image. Again the AOA is lesser in percentage and the areas outside are wider spread compared the PCA results.

Table 8: *Selected variables for FFS Raster Stacks.*

Raster Stack	Selected Variables
ffs_ind_only	"VVI" "SI" "TGI" "BI" "NGRDI"
ffs_ind	"VVI" "SI" "TGI" "green" "red"
ffs_ind_all	"BI" "HI" "RI" "VVI" "NDTI"
ffs_filtcor	"green_sobel9" "red_modal9" "green"
ffs_pca	"PC2" "red_modal9" "blue_min5" "blue_modal5"

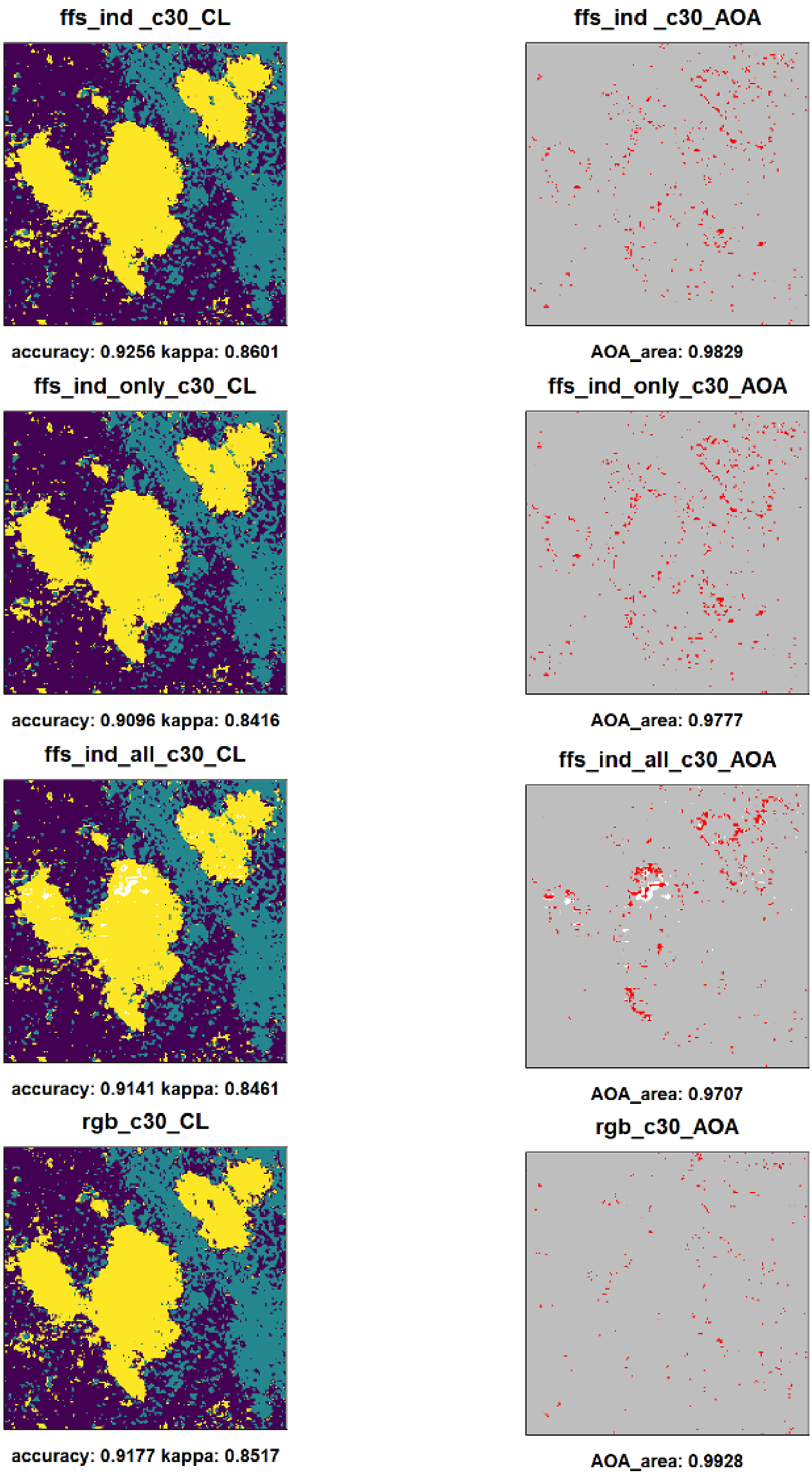

Fig. 9: *Prediction and AOA for test-series "*Impact of FFS*". Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red). Bottom: Results for RGB bands only for comparison.*

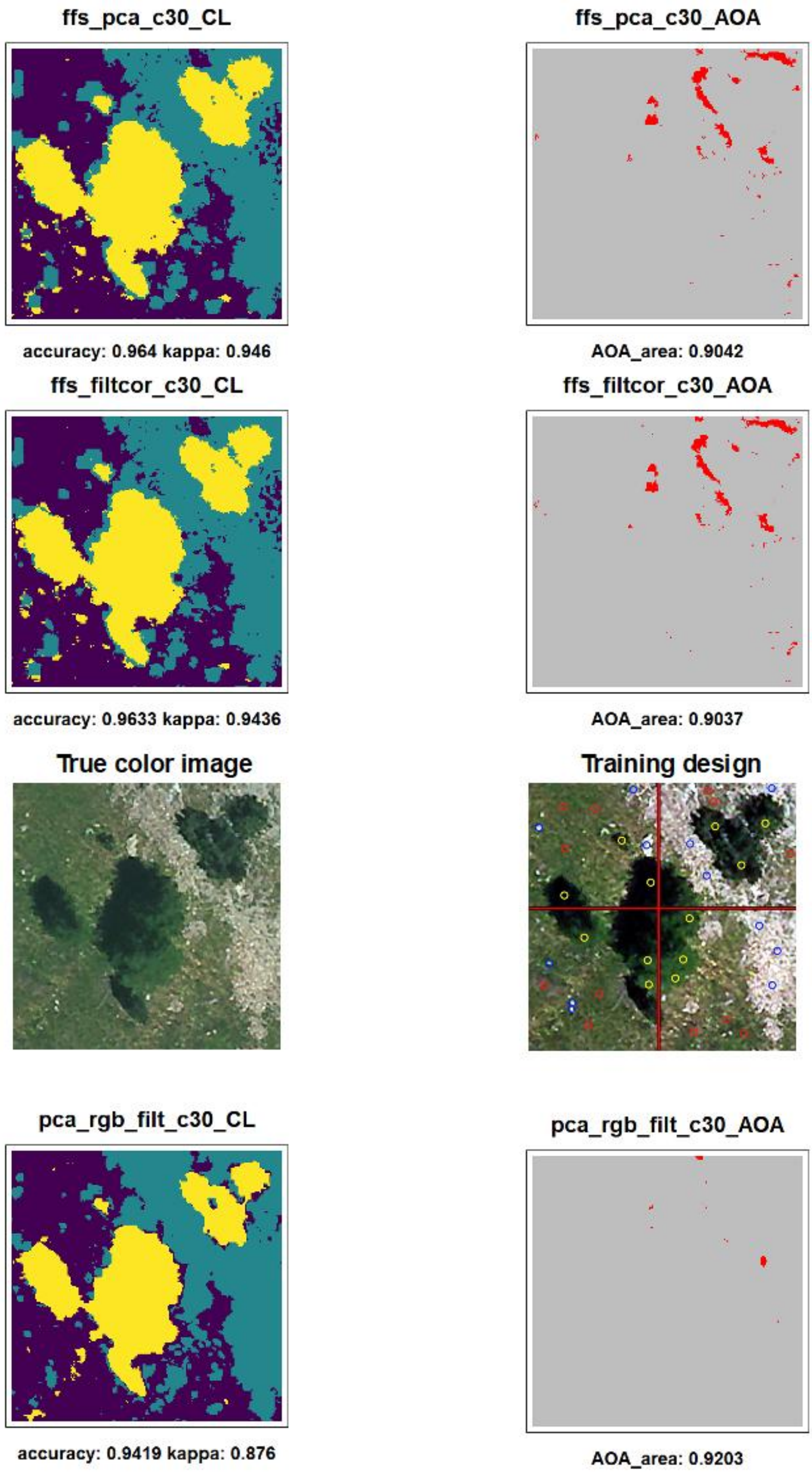

Fig. 10: *Prediction and AOA for test-series "*Impact of PCA*". Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red). Lower mid: True color image and Training design. Bottom: Results for 'pca_rgb_filt' for comparison.*

3.5 Fine Tuning for Best Result

The visual evaluated best results for the predictions are the FFS method for the 'ffs_filtcor' and 'ffs_pca' Raster Stacks as well as the PCA for the 'pca_rgb_filt' Raster Stack.

Overall the 'pca_rgb_filt' results seems to be more useful due to the lesser spread areas out of the AOA an its higher total amount.

The fine tuning results (see Fig. 11) for 'pca_rgb_filt' leads to a significant increase in model accuracy (up to 97.19) as well as visual evaluated performance for the prediction using the c15 instead of the c30 training design. While there are still lesser grass areas in the north east sector than seen on the original image this miss-classification is decreased compared to the results for the c30 training design. On the other hand the spots of small accumulations of class tree outside the real trees seem to be slightly greater with the c15 design. Like for the first test series with RGB bands only the designs with radius of 0.6 and 1.2 meter lead to a decreasing accuracy for the model as well as to decreasing performance overall with increasing areas of stone especially located with the tree shadow areas.

The AOA for the c15 design slightly decreases in total amount (91.66) compared to the c30 design (92.03) and the areas outside of the AOA are spread wider. Both designs c60 and c120 have lesser AOA with a significant distribution around the trees for the pixels outside the AOA.

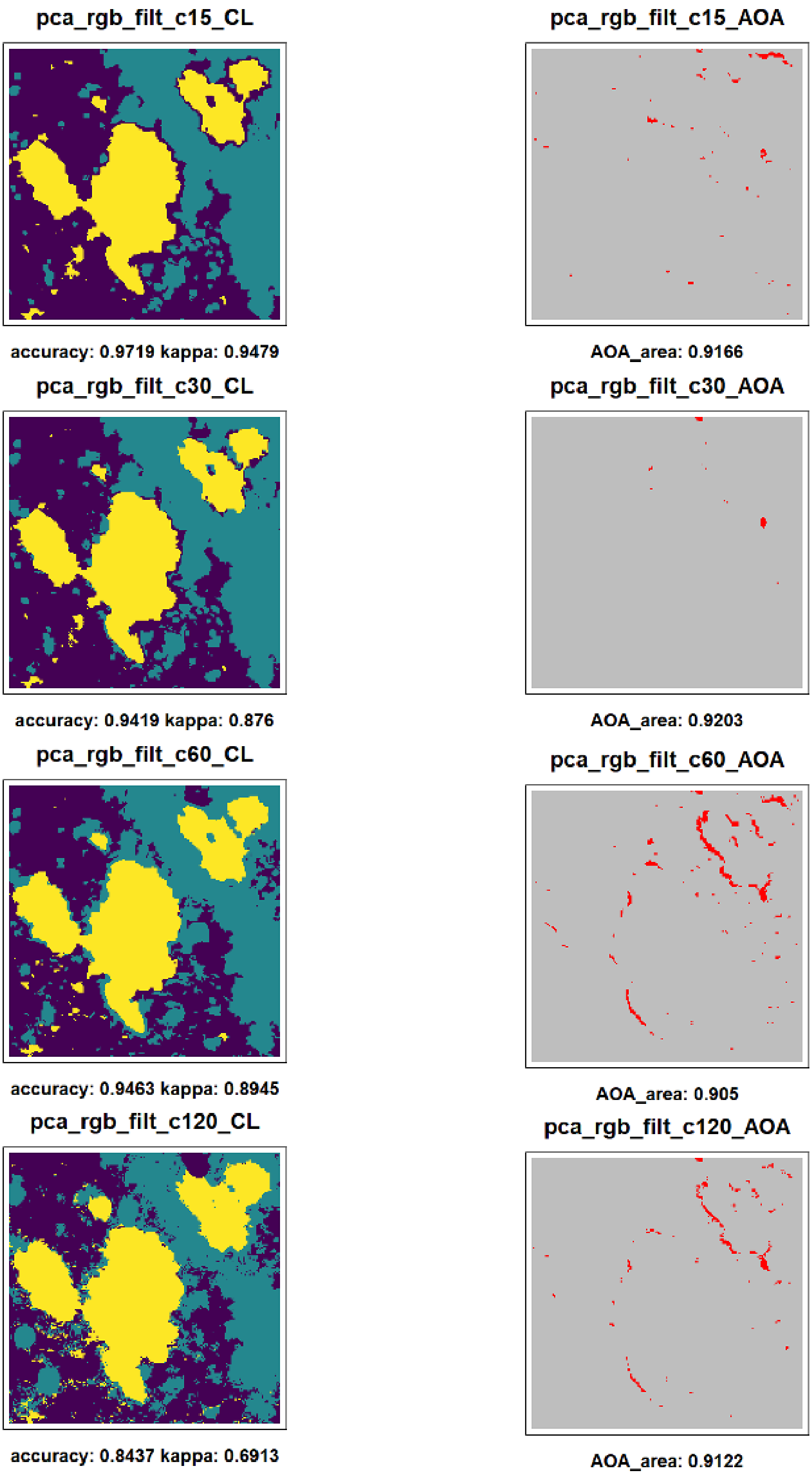

Fig. 11: *Prediction and AOA for test-series "*Fine Tuning (pca_rgb_filt)*". Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red).*

4. Discussion

Each of the dimensions has an influence on the results in different intensity. While some Data sets lead to mostly equal results others have significant impact on the prediction and the AOA. Further the impact on the AOA seems to be independent of the impact on the predictions. As expected the model accuracy does not deliver a precise value to estimate if the prediction is accurate but can be used as an orientation and for comparison. Due to the very high resolution of 0.15 meter and the fine distribution of grass and stone I did not apply any response layers for automated accuracy estimation for the prediction. The visual evaluation is therefore highly recommended but could lead to overoptimistic estimations due to the subjectivity. Basically I orientated my visual evaluation on the distribution of the classes assuming clear borders to be more accurate that high amount of single pixel with different classes. Further I assign unnatural forms like rectangle cluster to be less accurate. There are three areas where the influence can be observed very easily: The north east sector where the stone or open soil is dominant but has although some grass areas. If the grass is missing or very tiny in area the result can be seen as inaccurate. The second area are the trees in the middle of the image surrounded by grass and some stone spots. If too many pixels outside the trees are classified as trees and there are too much stone class the prediction may be less usable. At least the areas of the tree-shadow in North West direction from the trees seem to be areas which are difficult to classify.

In general the test series for the training design show that increasing sizes of the training areas lead to a decreasing model accuracy and AOA. Overall the results for the fine tuning seems to conform this. I assume that especially the training positions for the class 'stone' cause this effect. Several of the positions located in the southern and eastern areas of the image are very small in size so that wider training areas would catch the surrounding grass pixels for training the stone class. This theory can be supported by the noticeable high amount of stone in the north east sector where grass should be expected. For the AOI the design with cycles with a radius of 0.3 meter seem to lead to the best results. The form of the training areas have only minimal influence on the performance but cycles seems to be more useful leading to slightly more rounded results where especially wider rectangle lead to squared forms and edges in the prediction. I assume that natural forms tend to be more rounded than squared (eg trees) so that cycle training areas are recommended for classifications of natural classes.

The results for the second test series show that the use of artificially layers can improve the results. While none of the results really seems to be accurate the amount of single tree pixels is significant lesser than only using RGB bands. I suggest this effect comes from the use of filter which leads to a smoothing effect. This seems to be helpful for high resolution images. On the other hand this seems to cause rectangle clusters and a dramatically decease in precise classification for very small areas like the stones in the southern sectors.

The use of the PCA has significant impact on both prediction and AOA. But this effect is highly depending on the set of layers which have been used. The PCA for the Raster Stacks (with the dimensional reduction method) lead to slightly higher AOA but far more pixels for class tress spread over the prediction compared to the results without the PCA. On the other hand the PCA for only RGB bands and indices lead to results highly equal to the use of only RGB bands without PCA. The addition of the original RGB bands to the PCA has nearly no effect. I assume that the indices are highly correlated with each other and with the RGB bands. Therefore the use of a PCA for the RGB bands and the indices does not improve the performance of the classification. Furthermore the PCA has a slightly impact on the AOA especially for greater training area designs. Using only RGB bands lead to AOA values of over 99% (for Training designs with radius of 0.3 and 0.6 meter) while with a PCA the AOA is decreased to less than 98% for the c60 design and minimal decreases for c30.

The most significant impact of the PCA occurs if used for the filter computed for RGB and without the indices. Beside a more accurate prediction the areas outside the AOA are here accumulated in far lesser clusters. This effect indicates that the filter deliver far more information compared to the use of indices. Interesting is that the filter which are computed with wider moving windows seems to have nearly no effect on the PCA due to missing huge clusters which could be caused by greater moving window sizes. This indicates that for RGB images indices are less useful while spatial filter combined with a PCA can improve the results compared to only using the original RGB bands.

Comparing the FFS results with those for the PCAs which have been computed from the same Raster Layers there is no significant impact on the resulting prediction but on the AOA. The predictions for both methods are visually very equal. But there is higher model accuracy for the FFS approach. For the 'ffs_ind' Stack both 'green' and 'red' are selected what would explain the visual high similarity of the results compared to the use of only RGB or the PCA on RGB bands and indices. If the RI and HI are used for the Stack both are selected. I assume this can be explained by the high homogeneity of the values which lead to high performance values in the model but lesser realistic predictions. This indicates that the use of both indices is not recommended due to their negative influence on both the FFS and other methods.

A far greater impact occurs if the FFS is used on the correlated Raster Stack containing the filter layer. The resulting prediction is even more accurate that the 'pca_rgb_filt' result but has a lesser AOA. The 'ffs_filtcor' Raster Stack has lesser layers compared to the 'pca_rgb_filt' due to the reduction by correlation test. This may explain the differences but was used to reduce processing time for the FFS. At least if the first three PCAs are added to the Raster Stack the second is selected. This indicates that it is recommended to test both FFS and PCA against each other than only using the FFS.

One major aim of this study was to develop a comprehensible and reproducible workflow to test models for a best resulting prediction with high AOA and to identify the impacts of several parameters on the results. I can conclude that the data processing using the functions from the 'LEGION' function bundle is capable of computing several common used artificially layers. Further I can conclude that the setup of dimensions which where tested all have impacts on the results. So the used workflow is capable to test a wide range of variable and is able to detect the influences of the dimensions for a given AOI. While some dimensions have a significant impact which lead to an improvment of the resulting predictions and AOA others have minimal or negative effects. The 'LEGION::detectRstCor' and 'LEGION::detectRstHmgy' functions where not able to drop all layers which have negative influences on the results. The ‘brute’ force approach by first computing all available ALs and dimensional reduction alone does not lead to useful predictions. Probably further and more complex combinations of methods could improve the results like FFS used on the ‘hmgy_selected’ Raster Stack or even a FFS for the ‘full’ Raster Stack which have not been tested due to long processing times (without access to high performance computers). At least the ‘IKARUS_dawn’ function provides an easy way of standardized and comparable testing for the Raster Stacks with significant less effort of scripting compared to running all single steps for each data set. The developed workflow provides the ability of computing diverse data sets with combinations of several dimensions and to analyses the impacts on the prediction and AOA.

5. Conclusions

The developed workflow for computing and testing data sets for best resulting prediction with respective high AOA is a useful tool for classification approaches. It provides the ability to test the impact of different dimensions on the results for a given AOI. All those results depend on the used test area. Probably the effects of the dimensions differ for other training areas. But I suggest that I was able to build up a standardized workflow to test those impacts and make it easier to find the best performing setup. For RGB images the use of indices seems to be less useful due to high correlations. Filter can significantly improve the performance but although decrease it if used without a dimensional reduction method. Both the FFS and the PCA methods lead to high model accuracy and accurate predictions but should be tested against each other. Further the training design has significant impact on the results and should be tested in several designs. I finally recommend to first test the data sets and use the training designs for fine tuning to highly reduce processing time if high performance computers are not available.

References

BREIMAN, L. (2001): Random forests. Mach. Learn. 45, 5–32

CAO, W., LI, B. & Y. ZHANG (2003): A remote sensing image fusion method based on PCA transform and wavelet packet transform. International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003. Nanjing, China.

COULSTON, J., BLINN, C., THOMAS, V. & R. WYNNE (2016): Approximating Prediction Uncertainty for Random Forest Regression Models Photogrammetric Engineering & Remote Sensing Volume 82, Issue 3, March 2016, Pages 189-197

CUTLER, D., EDWARDS JR, T., BEARD, K., CUTLER, A., HESS, K., GIBSON, J. & J. LAWLER (2007): Random Forests for Classification in Ecology. Ecology 88(11). 2783–2792.

DEMSAR, U., HARRIS, P., BRUNSDON,C.,FOTHERINGHAM, A. & S. MCLOONE (2013): Principal Component Analysis on Spatial Data: An Overview. In: Annals of the Association of American Geographers 103(1). 106-128.

ESTORNELL, J. MARTI-GAVILA, J. SEBASTIA, M. & J. MENGUAL (2013): Principal component analysis applied to remote sensing Modelling in Science Education and Learning Volumen 6 (2). No. 7.

FIERENS, F. & P. ROSIN (1994): Filtering Remote Sensing Data in the spatial and feature domains. SPIEVol.2315. Imageand SignalProcessingfor RemoteSensing. 472-482.

FOX, E., HILL, R., LEIBOWITZ, S., OLSEN, A., THORNBRUGH, D. & M. WEBER (2017): Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. Environ Monit Assess 189: 316.

HIJMANS, R. (2020): raster: Geographic Data Analysis and Modeling. R package version 3.3-13. https://CRAN.R-project.org/package=raster.

HUNT E.R., CAVIGELLI M., DAUGHTRY C. , MCMURTREY III J. & C. WALTHALL (2005): Evaluation of Digital Photography from Model Aircraft for Remote Sensing of Crop Biomass and Nitrogen Status. Precision Agriculture 6. 359–378.

HUNT, E.R., DORAISWAMY, P.C., MCMURTREY, J.E., DAUGHTRY, C.S., PERRY, E.M. & B. AKHMEDOV (2013): A visible band index for remote sensing leaf chlorophyll content at the canopy scale. Int. J. Appl. Earth Obs. Geoinf. 21, 103–112.

JIN, H., STEHMAN,S. & G. MOUNTRAKISA (2014): Assessing the impact of training sample selection on accuracy of an urban classification: a case study in Denver. International Journal of Remote Sensing. Vol. 35, No. 6, 2067–2081.

KUHN, M. & K. JOHNSON (2013): Applied Predictive Modeling, 1st ed. Springer, New York.

KUHN M. (2020): caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret.

LEUTNER B. , HORNING N. AND J. SCHWALB-WILLMANN (2019): RStoolbox: Tools for Remote Sensing Data Analysis. R package version 0.2.6. https://CRAN.R-project.org/package=RStoolbox.

MEYER, H. & E. PEBESMA (2020): Predicting into Unknown Space? Estimating the Area of Applicability of Spatial Prediction Models.

MEYER, H., REUDENBACH, C., WÖLLAUER,S. & T. NAUSS (2019): Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecological Modelling 411.108815.

MEYER, H., REUDENBACH, C., HENGL, T., KATURJI, M. & T. NAUSS (2018): Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9.

MEYER, H. (2020): CAST: 'caret' Applications for Spatial-Temporal Models. R package version 0.4.2. https://CRAN.R-project.org/package=CAST.

R CORE TEAM (2020): R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. Austria. URL https://www.R-project.org/.

RAY, S., SINGH, J., DAS, G. & S. PANIGRAHY (2004): Use of High Resolution Remote Sensing Data for Generating Sitespecific Soil Management Plan Conference: The Int. Arch. Photogram. Rem. Sens. & Spatial Inform. Syst At: Istanbul Volume: 35 (B7)

THE IDB PROJECT (2020): Index Database https://www.indexdatabase.de/

ZHAO, D., JIAO, Y., WANG, J., DING, Y. LIU, Z., LIU, C., QIU, Y., ZHANG, J., XU, Q. & C. WU (2020): Comparative performance assessment of landslide susceptibility models with presence-only, presence-absence, and pseudo-absence data. Journal of Mountain Science 17(12).