**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). A common used algorithm is Random Forest (Breiman 2001) which is known for relatively high stability and robustness. Random forest is a non-parametric ensemble approach and suffers a direct quantification of prediction error (Coulston et al 2016). A popular way for estimating the performance of a model is the k-fold cross validation (CV) approach (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. Especially for low spectral resolution data sources like RGB images the performance of the model can be increased by giving the algorithm more value information due to data processing (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). Machine learning algorithms in general tend to be sensibility to over-fitting due to over-sized data sets, high correlating data or layer which contains continuous data (like coordinates) (Meyer et al 2016). So a dimensional reduction is highly recommended. A forward feature selection (FFS) (Meyer et al 2019) can be used to determine the best fitting layers but although can result in long processing times. 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. This is very important especially for small scale classifications of objects on high resolution images.

I this study I will test the impact of the training design on the prediction and AOA for a small scale classification of trees, grass and soil for an high resolution RGB image at the upper alpine treeline. I assume that very small training areas would deliver less information and result in inaccurate predictions but greater training areas would catch to many values from other classes which would reduce the performance too. So I hypothesize that increasing sizes of training areas will first result in increased performances for the predictions but decreasing performance after a maximum is reached. With the AOA highly depending on the values and its distribution I assume that more data could result in lesser AOA. Therefore I hypothesizes that increasing sizes for the training areas lead to decreasing AOA.

Besides the original RGB data I will compute a Raster Stack of several artificially layers to compare the performance and processing time. The Raster Stack will include spatial filter, spectral indices and PCA. For dimensional reduction I will drop highly correlation layers and use a FFS to select the best fitting layers. I assume that the performance will be higher compared to the use of only RGB bands but would take more time to proceed.

For validation strategy I will use three sets of training points estimating that each set will have the highest performance with the same training design and further have only minimal differences compared to each other.

**2. Data and Methods**

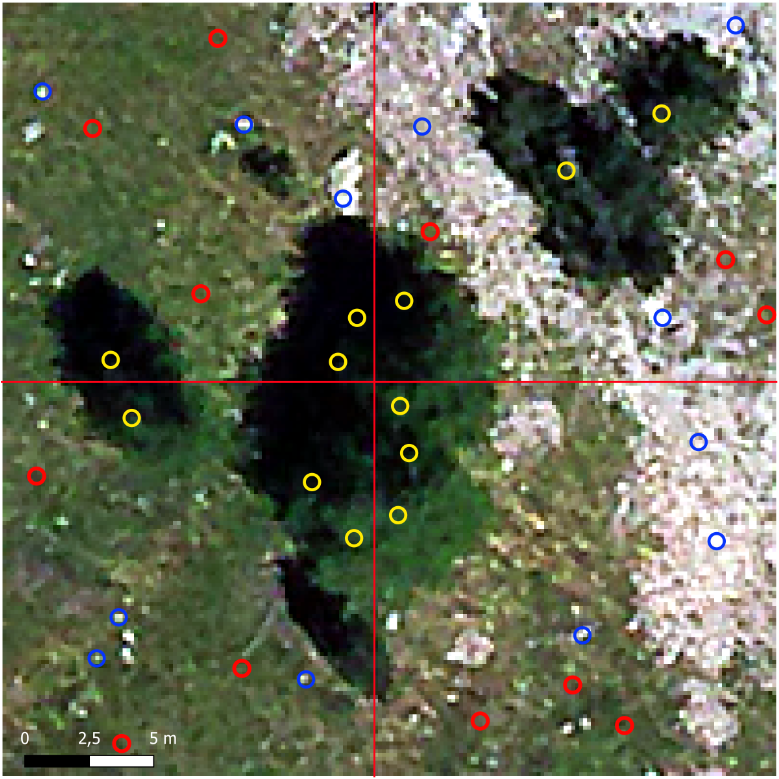
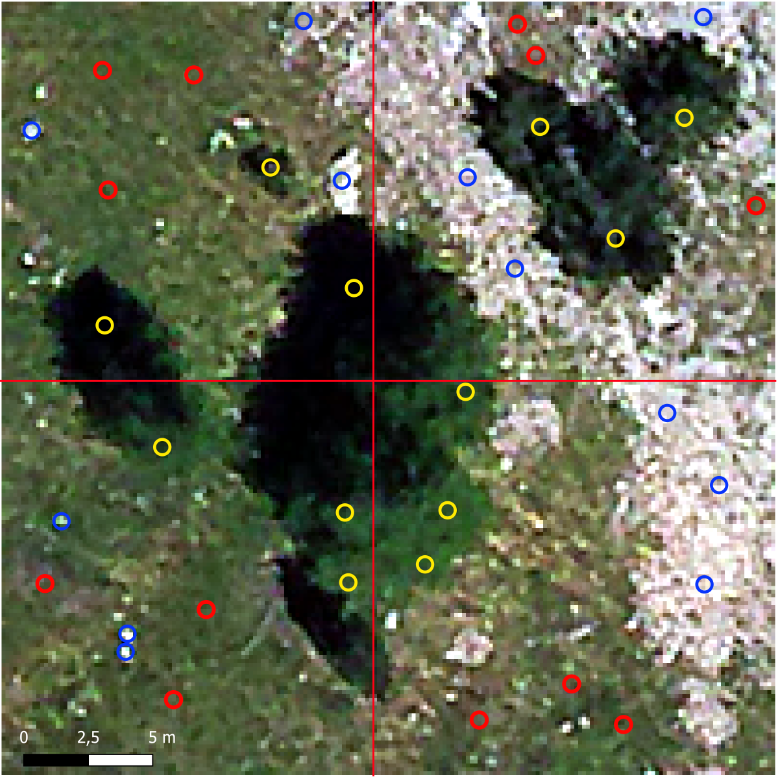
***Fig.1 :*** *Schematic “map” of the multidimensional hyperspace: Each Dimension has multiple to infinite variables and depends on others (blue arrow) and increases or multiplies the amount of resulting variables in the hyperspace (red arrow). Each processing step (black arrow) merges individual hyperspaces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

For this study I will use a high resolution RGB image and compute several artifically layers The data processing and modelling is performed in R version 4.0.4 (R Core Team 2020). The used Image, Training Point Layer 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 (upper left) and training points set 1 (upper right), set 2 (lower left), set 3 ((lower right). The training points: yellow= tree, blue= soil, red=grass.*



For the study area I will use a 30x30 meter aerial image with 0.15 meter spatial resolution (see fig.2). The scene shows some trees surrounded by grass and open soil located at the upper alpine treeline in the Lautaret vally (French alps). For the classification I will use 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 and training design. For a more standardized and comparable testing I will use points for the training. 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 (Fig. 1). The points are set on positions representing the average behavior of the classes. 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. I will use spatial buffers around the points with same ranges for each point to compute different training designs in forms of both cycles (c) and rectangles (r).

2.2 Artificially Layer Computation

Besides the original RGB bands I will use a Raster Stack with artificially layers (Tab. ). 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 (grass/trees) due the sensitivity to chlorophyll ( Hunt et al 2005; Hunt et al 2013).. 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. With the extents of the Study area and the 0.15 meter resolution I will use Moving Windows of 3x3, 5x5, 7x7 and 9x9. For the computation of the artificially layers I will use common spatial filter based on the 'Raster::focal' function (Hijmans 2020) and sobel filter for edge detection as well as common used spectral RGB indices (Table 2). For further information about the spectral RGB indices see: The IDB Project (2020); Ray et al (2004); Hunt et al (2013). Another strategy to generate information for the machine learning 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). The first PCA can be a value predictor regarding to Meyer et al (2019) if computed for the RGB bands and the indices. I will use the 'rasterPCA' function provided by the 'Rstoolbox' package (Leutner at al 2019). From this i will first compute the RGB indices and a PCA for the RGB bands and the Indices. Next I will compute all filter with Moving Windows of 3x3, 5x5, 7x7 and 9x9 for the first PCA.

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. Besides a FFS would be able to select the best fitting layers too many layers would result in long processing. 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. For the correlation test I will drop all layer with a cor value of >= 0.9 and <= -0.9 and further drop homogeneous layers which have >= 90% of data values in <=10% of the data range

***Table 1:*** *Artificially layersused for the Raster Stack*

|  |  |  |  |
| --- | --- | --- | --- |
| Index 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) |
| Filter 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 | | |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |

2.6 Test Method

Machine Learning

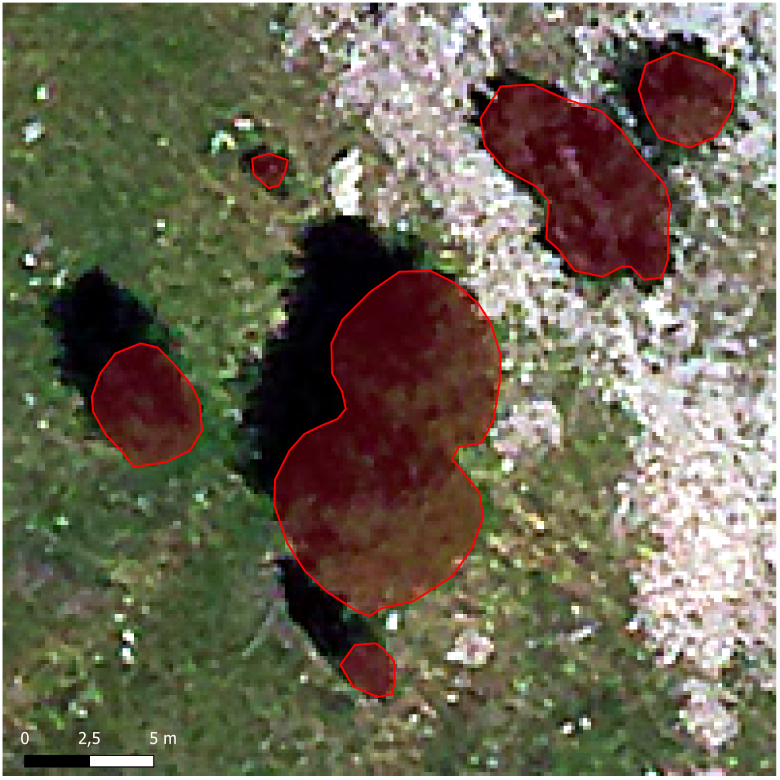
For the computation of the model and the respective AOA calculation I will use a workflow mainly based on 'CAST' (Meyer 2020). First the training data is extracted for the desired training design. Next the LLOCV is prepared using 'CAST::CreateSpacetimeFolds' (Meyer 2020) by defining indices for the data to be left out based on the four sectors of the study area. To prevent over-fitting I will use a forward feature selection (FFS) implemented in ‘CAST’. The 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). For the machine learning algorithm the ‘Random forest’ (Breiman 2001) is used. The resulting model based on the selected variables will be used to predict the classification for the study area. At least the AOA is calculated using 'CAST::aoa' (Meyer 2020) . For the resulting performance for the model the ‘accuracy’ and ‘kappa’ will be used as well as the percentage of AOA. Further the runtime for the process is saved for comparison reasons.

Validation Concept

For validating the resulting prediction I will use a response layer for the tree class. The response layer is digitalized by hand representing the visual estimated areas of trees in the study area. This layer is compared to the areas which are classified by the algorithm as class tree. Estimating the response layer to be accurate the performance of the prediction will be estimated by the amount of cells classified as trees which overlap with the response layer (‘hitrate’) and those outside the responds areas (‘missrate’). The performance of the prediction will be estimated by the ‘accuracy’ and ‘kappa’ values as well as the response values of ‘hitrate’ and ‘missrate’. The AOA is primarily evaluated by its total percentage but further visually to check for significant spatial pattern. To validate the results I will use the method for all three training point sets and compare the results.

Test Series

To investigate my hypothesis I will use increasing training areas in cycle and rectangle forms for both RGB bands and the computed Raster Stack. The training areas will be computed by using spatial buffer with a given radius around the training points. Due to the resolution and the extent of the image I estimate radius of 0.15, 0.30, 0.45, 0.60 and 0.75 to be useful to test. training assume over XX radius would wider than the objects which should be classified and easily catch values for more than one class which would lead to false results.(GGF NACHBEREITEN).



**3 Results**

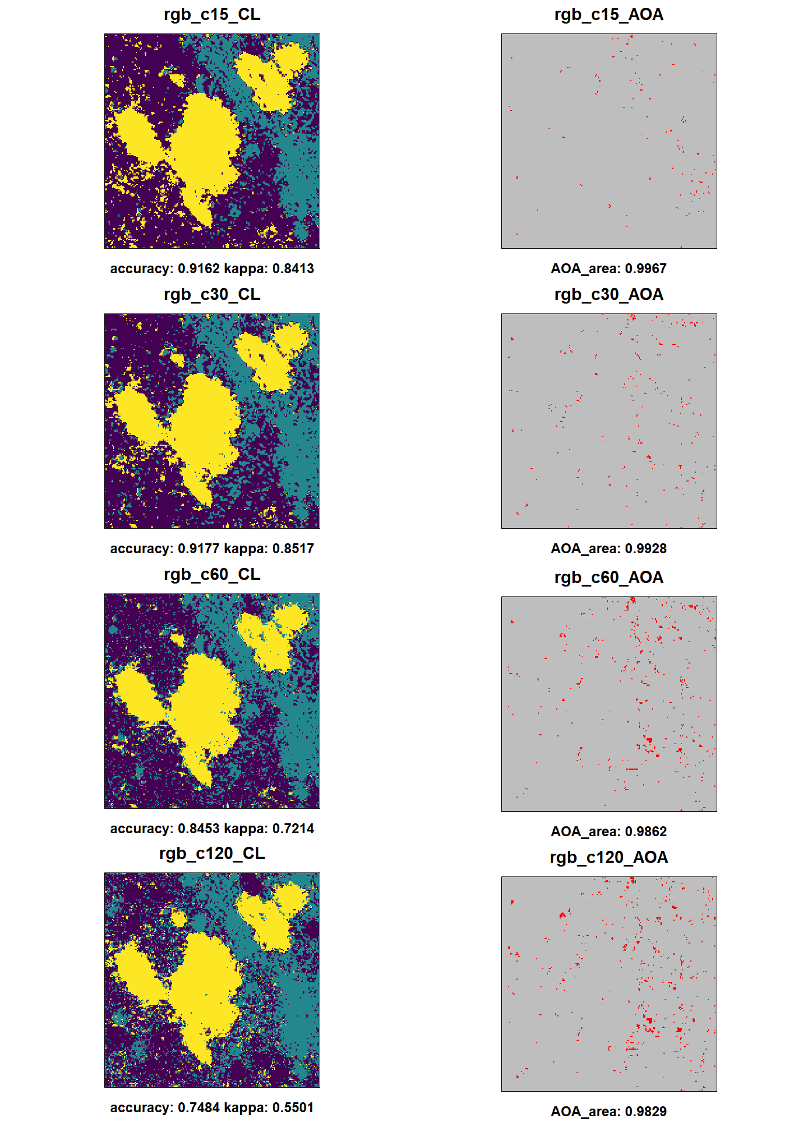
In total I have tested 28 combinations of Raster Stack and training design. For the prediction the accuracy and kappa of the model is printed as well as the percentage of the AOA.

3.1 Impact of Training Design

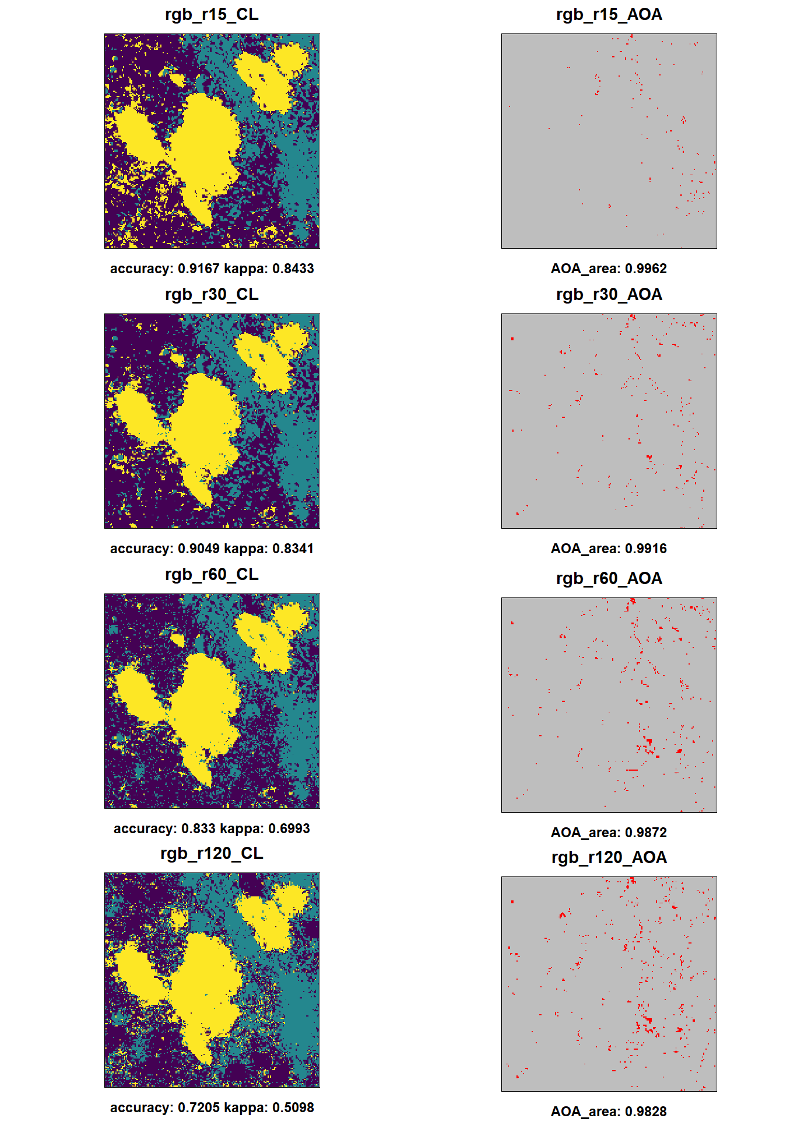
There is only a slightly difference between the cycles (Fig. 4) and rectangle (Fig. 5) designs despite for which sizes. Overall the rectangles have minimal lesser accuracy compared to the cycle forms except for the 0.15 meter sizes where the rectangles are slightly higher in performance. The performance 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.

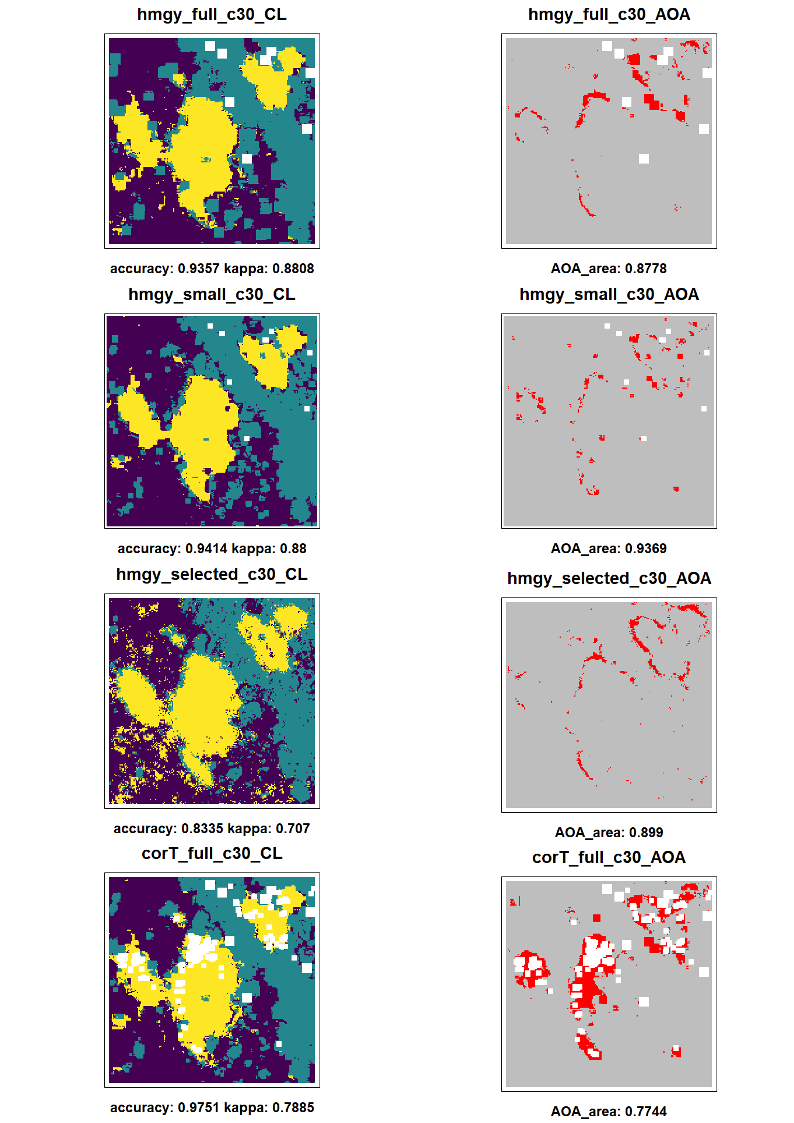


***Fig. 4:*** *Prediction and AOA for test-series ”*Impact of training design (cycles)*”. Left: Prediction with classes: Trees: yellow, grass: dark blue, stone/open soil: blue-green. Right: Cells in AOA (grey) and outside AOA (red).*



***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).*

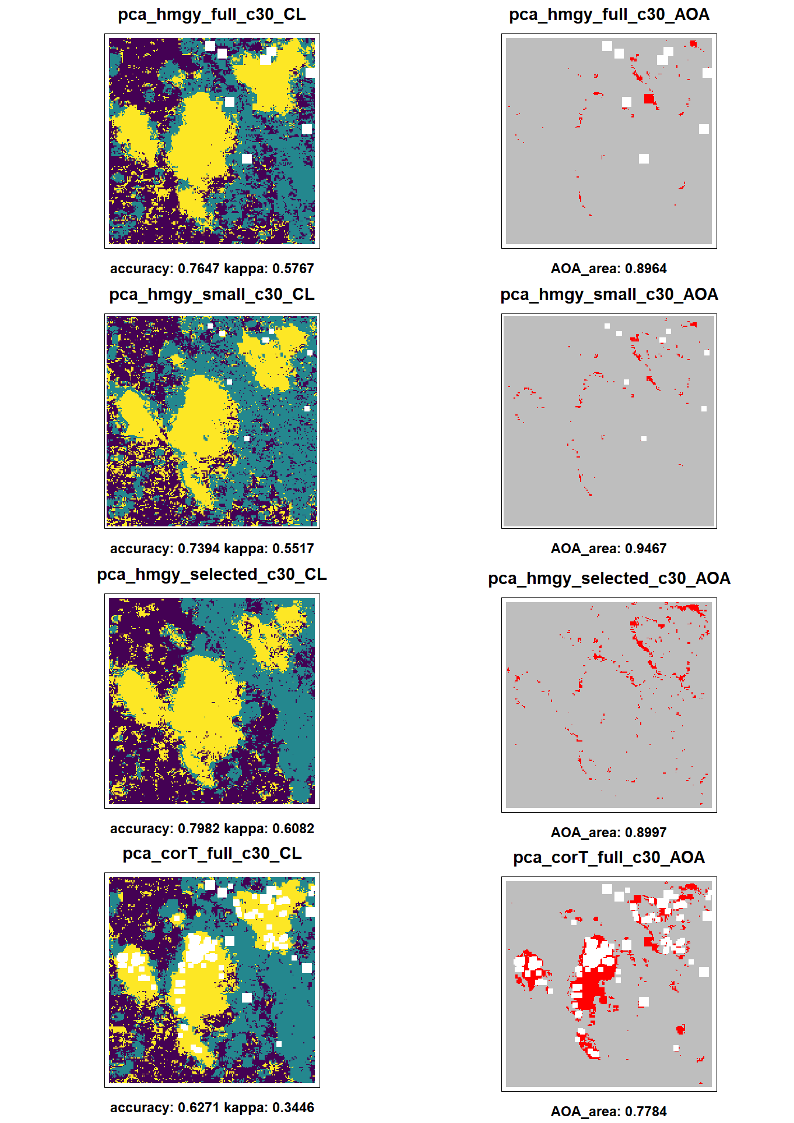


3.3 Impact of PCA

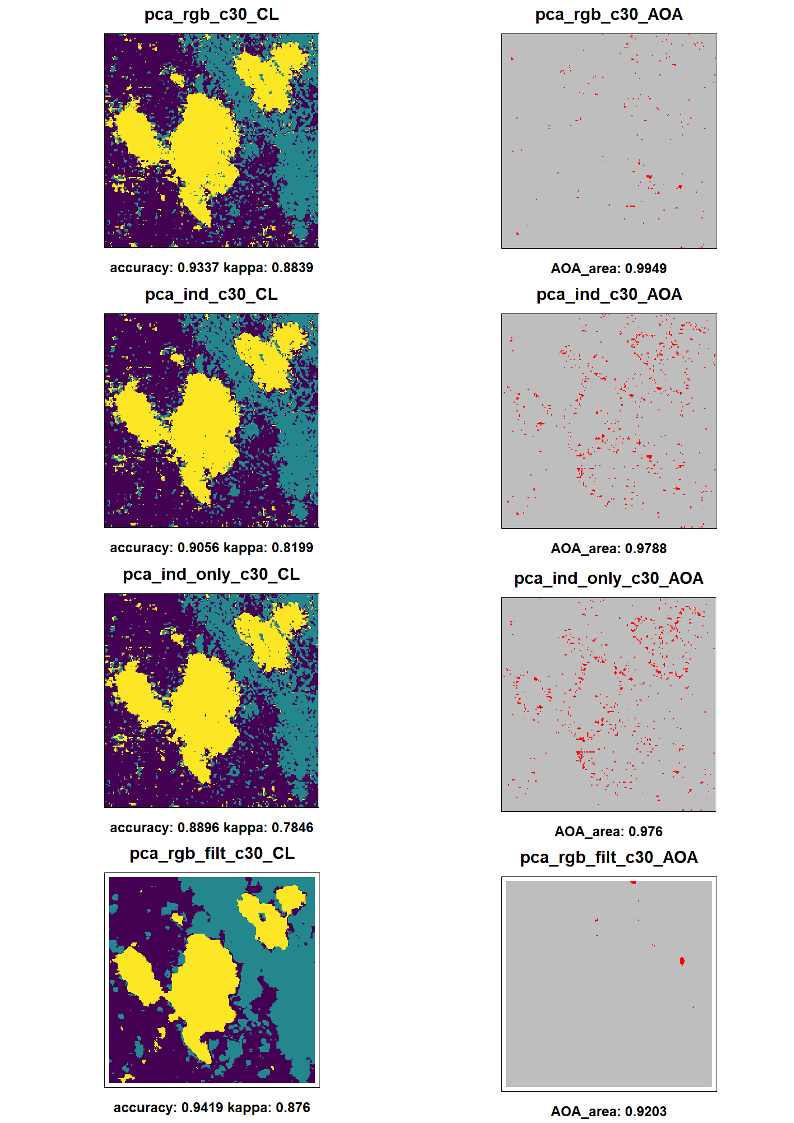
The results for the PCAs computed from the artificially layer Raster Stacks lead to significant lesser accuracy (Fig. 7) compared to the models using the respective Raster Stack (Fig. 4 ) decreasing from almost around 0.90 to 0.70. Further the visual evaluation shows a far lesser separation of the classes. 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).*



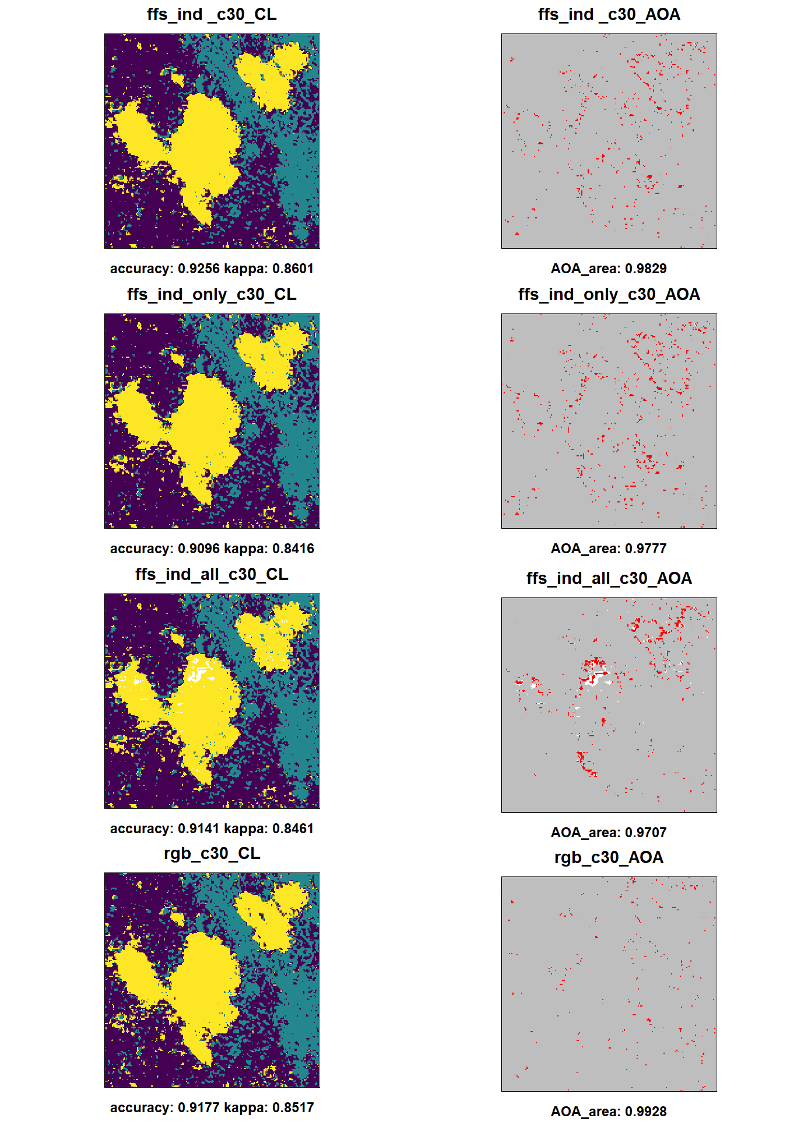
3.4 Impact of FFS

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.

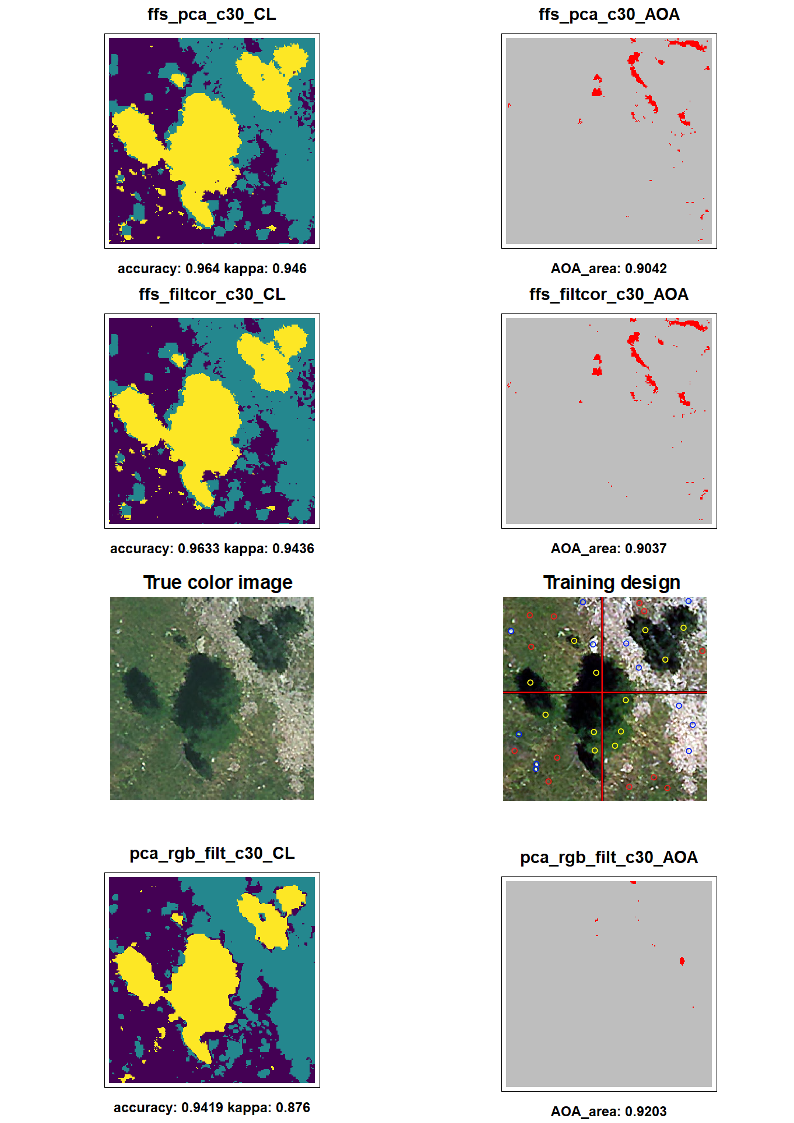
***Table 8:*** *Selected variables for FFS Raster Stacks.*

|  |  |
| --- | --- |
| 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

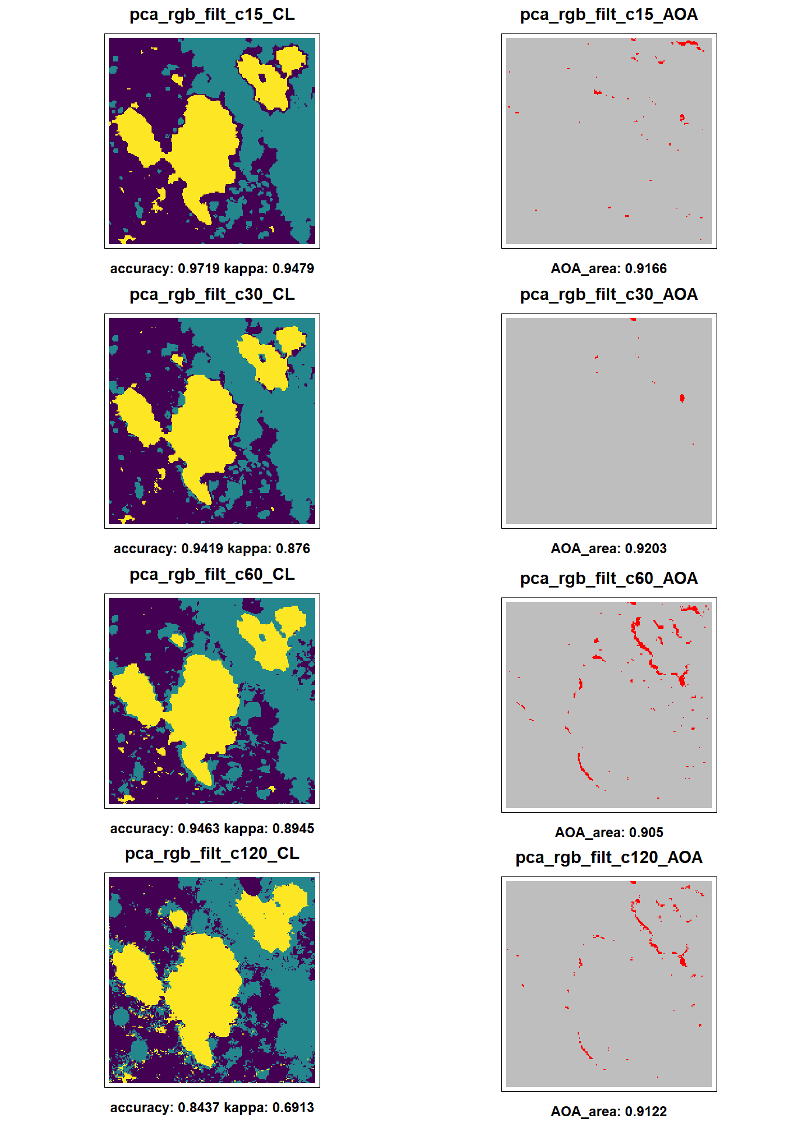
The visual evaluated best results for the predictions are the FFS method for the 'ffs\_filtcor' and 'ffs\_pca' Raster Stacks as well as the PCA for the 'pca\_rgb\_filt' Raster Stack.

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](https://ieeexplore.ieee.org/xpl/conhome/9008/proceeding). 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).