

CHARLES UNIVERSITY OF PRAGUE

Faculty of Sciences

Study programme: Geography

Study specialization: Geoinformatics and Cartography



Bc. Lucie HROMÁDKOVÁ

Classification of meadow vegetation in the Krkonoše Mts. using aerial hyperspectral data and support vector machines classifier

MASTER THESIS

Supervisor: RNDr. Lucie Kupková, Ph.D.

Consultant: RNDr Stanislav Březina, Ph.D.

Prague, 2015

Zadání diplomové práce

pro Lucii Hromádkovou
obor Kartografie a geoinformatika

Název tématu:

Classification of meadow vegetation in the Krkonoše Mts. using aerial hyperspectral data and support vector machines classifier

Zásady pro vypracování

Meadow vegetation in The Krkonoše Mountains is very complex and heterogeneous environment with plenty of diverse vegetation communities. As this heterogeneity makes it difficult to use conventional classifiers successfully, non-conventional classifier - support vector machines (SVM) was chosen for this task. The thesis will focus on classification setup proposal of meadow vegetation in The Krkonoše Mountains using aerial hyperspectral data achieving as accurate classification as possible given the available data, methods and user requirements.

The first goal of classification will be to experimentally determine which kernel and specific values set for kernel parameters and also value of penalty parameter give the highest classification accuracy for the area of interest.

Commonly, the result (accuracy) of classification is influenced by selection of training pixels and their sampling design, such selection is of even higher importance when complex environment is being classified. The size of training dataset is on the other hand linked to classifier to be used and also to features of dataset to be classified. Unlike conventional classifiers, SVM classifier, as a boundary classifier, uses only pixels on the boundary of classes, because these pixels are the most informative for placing hyperplane which

separates the classes. Thus the minimum size of training set is assumed to be smaller for SVM classifier than for conventional classifiers while the accuracy of classification stays high (Foody and Mathur, 2004).

Based on this hypothesis of Foody and Mathur we will experiment with size of training dataset for particular classes and also with distribution of training polygons in the imagery, related to the accuracy of classification. Based on this experiment **the second goal of the thesis will be to propose ideal design for training dataset** e. g. to propose 1) minimum size of the training set for each class (number of polygons, number of pixels), 2) ideal spatial distribution of training dataset (regular vs. irregular), 3) if for more abundant classes in the imagery more training pixels should be selected and how clustered they should be, 4) if there are some differences between different legend classes. The criterion of tests will be the result of classification accuracy. Results of these experiments could be also used in future as guideline for field mapping, to limit extensive whole-area mapping to only “hot spots” shown as possible sources of support vectors.

An additional goal of the thesis will be classification of the same training dataset using neural network method. This additional goal will be performed mostly in order to compare the result (accuracy) of classification using SVM to a result obtained by another, known, method. Also this experiment will either confirm or reject the hypothesis of Foody and Mathur, that SVM can achieve the same classification accuracy as other classifiers however using smaller training datasets (Foody and Mathur, 2004).

Rozsah grafických prací: dle potřeby

Rozsah průvodní zprávy: 50-60 stran textu, mapové přílohy – výstupy klasifikací

Seznam odborné literatury:

Belousov A.I., Verzakov S.A, von Frese J., 2002, A flexible classification approach with optimal generalisation performance: support vector machines; chemometrics and Intelligent laboratory systems, 64. 15 – 25;

Benediktsson J.A, Swain P.H., Ersoy O.K; 1990, Neural network approaches versus statistical methods in classification of multisource remote sensing data; IEEE Transactions on geoscience and remote sensing, 28/4, 540 – 551;

Borengasser M., Hungate W.S., Watkins R.; Hyperspectral remote sensing: principles and applications; Taylor&Francis, 2008;

Camps-Valls, G., Gómez-Chova L., Calpe-Maravilla J., 2004, Robust support vector method for hyperspectral data classification and knowledge discovery, IEEE Transactions on geoscience and remote sensing, 20, 1 – 13;

Chan J.C-W., Beckers P., Spanhove T., Borre J.V.; 2012, An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery; International Journal of Applied Earth Observation and Geoinformation, 18, 13 – 22;

Foody G.M., Mathur, A., 2004; Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification, Remote sensing of environment, 93, 107 – 117;

Huang, C., Davis, L.S., Townshend, J.R.G, 2002, An assessment of support vector machines for land cover classification; International journal of remote sensing, 23/4; 725 – 749;

Jones H.G. Vaughan R.A.; Remote sensing of vegetation: Principles, techniques and applications; Oxford: Oxford university press, 2010;

Marcinowska, A., Zagajewski, B., Ochtyra, A., Jarocińska, A., Raczko, E., Kupková, L., Štych, P., Meuleman, K., 2014, Mapping vegetation communities of the Karkonosze National Park using APEX hyperspectral data and Support Vector Machines; Miscellanea Geographica – Regional Studies on Development; 18/2, 23-29;

Melgani F., Bruzzone, L., 2004; Classification of Hyperspectral Remote sensing images with support vector machines; IEEE transactions on geoscience and remote sensing, 42/8; 1778 – 1790;

Mountrakis G., Im J., Ogole C., 2011; Support vector machines in remote sensing: A review; ISPRS Journal of Photogrammetry and Remote Sensing, 66, 247 – 259;

Pal M. and Mather P.M., 2005; Support vector machines for classification in remote sensing, International Journal of Remote Sensing, 26/5, 1007 – 1011;

Vedoucí diplomové práce: RNDr. Lucie Kupková Ph.D.

Konzultant diplomové práce: Mgr. Stanislav Březina, Ph.D.

Datum zadání diplomové práce: 30.1.2014

Termín odevzdání diplomové práce: červenec 2015

Platnost tohoto zadání je po dobu jednoho akademického roku.

.....
Vedoucí diplomové práce

.....
Vedoucí katedry

V Praze dne 29.1.2015

Acknowledgement:

I would like to express my gratitude to my supervisor RNDr.Lucie Kupková, Ph.D. for the useful comments, remarks and engagement through the learning process of this master thesis. Furthermore I would like to thank Mgr.Stanislav Březina, Ph.D. for valuable information about vegetation communities. Also, I would like to thank Mgr.Lucie Červená and RNDr.Jakub Lysák for their precious time and shared advice. Last but not least, I would like to thank my loved ones, who have supported me throughout the entire process.

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Trutnově, 24.7.2015

.....

Lucie Hromádková

ABSTRACT

Meadow vegetation in the Krkonoše Mountains National Park is classified in this master thesis using aerial hyperspectral data from sensor AISA and Support Vector Machines (SVM) and Neural Networks (NN) classification algorithms.

The main goals of the master thesis are to determine the best settings of SVM parameters and to propose an ideal design for a training dataset for this classification algorithm and mapping of the meadows in the Krkonoše mountains. The criterion of the tests will be the result of classification accuracy (confusion matrices and kappa coefficient). The additional goal of the master thesis is to compare performances of both utilized classifiers, especially regarding the amount of training pixels necessary for successful classification of the mountainous meadow vegetation.

Classification maps of the area of interest and Python scripts are the main outputs of the master thesis. These outputs will be handed over to the Administration of the Krkonoše Mountains National Park for further utilization in the monitoring and protecting these valuable meadow vegetation communities.

Key words: hyperspectral data, AISA, Support Vector Machines, Neural Networks, training dataset, mountainous meadow vegetation

ABSTRAKT

Tato diplomová práce se zabývá využitím leteckých hyperspektrálních dat senzoru AISA a klasifikačních metod Support Vector Machines (SVM) a Neural Networks (NN) pro mapování lučních společenstev v Krkonošském národním parku.

Hlavní cíle práce jsou experimentální určení nejlepší kombinace parametrů algoritmu SVM, a návržení ideálního trénovacího datasetu pro tento algoritmus a krkonošská luční společenstva. Kritériem úspěšnosti jednotlivých kombinací parametrů SVM a trénovacích datasetů jsou výsledky posouzení přesnosti klasifikace pomocí confusion matic a kappa koeficientu. Kromě hlavních cílů je účelem práce také porovnání klasifikačních algoritmů SVM a NN, především co se týče počtu trénovacích pixelů potřebných pro úspěšnou klasifikaci horských luk.

Hlavními výstupy práce jsou klasifikační mapy zájmových území a skripty v jazyce Python, které budou předány Správě KRNAP pro další využití v monitoringu a ochraně cenných lučních společenstev.

Klíčová slova: hyperspektrální data, AISA, Support Vector Machines, Neural Networks, trénovací dataset, horská luční vegetace

Table of contents

Table of contents	5
List of abbreviations (in alphabetical order)	7
List of figures	8
List of tables	9
1. Introduction	10
2. Theoretical background.....	12
2.1 Remote sensing of vegetation	12
2.1.1 Spectral properties of a leaf.....	12
2.1.2 Spectral behaviour of canopies – “hotspot” phenomenon	15
2.2 Hyperspectral data in vegetation mapping	18
2.2.1 Methods for reducing the dimensionality of hyperspectral data	19
2.3 Classification methods currently used for grassland vegetation mapping	21
2.3.1 Probabilistic approach	21
2.3.2 Gradient modelling.....	22
2.3.3 Vegetation indices	23
2.3.4 Machine learning algorithms.....	24
2.3.5 Object-oriented classification.....	29
2.4 Methods used for the vegetation classification in the Krkonoše Mountains.....	30
3. The study area	33
3.1 Climate	33
3.2 Classified types of biotopes.....	35
3.2.1 Legend.....	35
3.2.2 Description of the classified biotopes	36
4. Data	45
4.1 Aerial imagery.....	45
4.1.1 The description of utilized AISA sensors	46
4.1.2 Pre-processing of aerial imagery	48
4.1.3 Errors in the aerial data	52
4.2 Field mapping.....	54
4.2.1 Pre-processing of the field data	54
4.2.2 Problems with the field data.....	56
5. Methods.....	59
5.1 Tuning parameters of the SVM classifier (step 1)	60
5.1.1 The penalty parameter	61
5.1.2 The choice of a kernel	62

5.2 Defining a training and a validating dataset (step 2).....	63
5.2.1 The amount of training pixels	63
5.2.2 Sampling design	65
5.3 ANN setup.....	69
5.4 Assessment of the accuracy of classification	70
6. Results	72
6.1 The SVM classifier.....	72
6.1.1. The parameters of the SVM classifier	72
6.1.2 Sampling design and the amount of training pixels	74
6.1.3 The accuracy assessment of the classification of the enclaves of interest	75
6.2 The ANN classifier.....	78
6.2.1 The parameters of the ANN classifier	78
6.2.2 The accuracy assessment of the classification of the enclaves of interest	79
6.3 Classification maps for the selected cases.....	82
7. Discussion	87
7.1 The utilized datasets and their limitations.....	87
7.2 The tested parameters of the SVM classifier	89
7.3 Sampling design and the amount of training pixels	90
7.4 Results of classification.....	93
7.5 A comparison of the results of classification with results from other studies.....	96
8. Conclusion.....	98
9. References	101
10. Appendices	101

List of abbreviations (in alphabetical order)

AISA	Airborne Imaging Spectrometer for Applications
ANN (NN)	Artificial Neural Networks (n eural n etworks)
AOPKČR	Agentura ochrany přírody a krajiny České republiky (Nature Conservation Agency of the Czech Republic)
APEX	Airborne Prism Experiment
ASD	Analytical Spectral Devices
AVIRIS	Airborne Visible / Infrared Imaging Spectrometer
BR	bidirectional reflectance
BRDF	bidirectional reflectance distribution function
CCD	charge-coupled devices
CE	commission error
ČÚZK	Český úřad zeměměřičský a katastrální (State Administration of Land Surveying and Cadastre)
DAIS	A Digital Airborne Imaging System
DEM	digital elevation model
DSM	digital surface model
DT	decision tree
ENVI	Environment for Visualizing Images
ESA	The European Space Agency
ESR	equal sample rate
ESS	equal sample size
FOV	field of view
GCP	ground control point
GPS	global positioning system
HyMap	Hyperspectral Mapper
IDL	Interactive Data Language
IMU	inertial measurement unit
IR	infrared radiation
KRNAP	Krkonošský Národní Park (The Krkonoše Mountains National Park)
Landsat ETM+	Landsat Enhanced Thematic Mapper Plus
Landsat TM	Landsat Thematic Mapper
LAI	leaf area index
LSU	linear spectral unmixing
MLC	maximum likelihood classifier
MLP	multilayer perceptron; type of ANN
MNF	minimum noise fraction transformation
MCT	mercury – cadmium – telluride
NDVI	normalized difference vegetation index
NIR	near infrared region of spectrum
OA	overall accuracy
OBIA	Object-Based Image Analysis
OE	omission error
PA	producer's accuracy
PCA	principal component analysis
RBF	radial basis function; the type of ANN activation function or type of kernel in SVM
SAM	spectral angle mapper
SAVI	soil adjusted vegetation index
SNR	signal to noise ratio
SVM	Support Vector Machines
SWIR	short-wave infrared region of spectrum
TR	training rate
TTC	training threshold contribution
UA	user's accuracy
VI	vegetation index / vegetation indices
VNIR	visible and near infrared region of spectrum

List of figures

Figure 1	Spectral features of key components of leaves	12
Figure 2	A spectral curve of a leaf	13
Figure 3	An aerial image of a vineyard from a balloon	15
Figure 4	A hotspot diagram	16
Figure 5	A hyperspectral cube	17
Figure 6	Schematics of the new dataset found by PCA	18
Figure 7	A schema of a neural networks setup	24
Figure 8	An example of linear SVM	26
Figure 9	The transformation of input space into higher dimensional feature space and constructing a separating hyperplane	27
Figure 10	A map of the classified enclaves	33
Figure 11	Closed alpine grasslands dominated by <i>Nardus stricta</i>	36
Figure 12	Oligotrophic grasslands	37
Figure 13	A montane <i>Trisetum</i> meadow	38
Figure 14	Waterlogged grasslands	39
Figure 15	A locally waterlogged area with <i>Myosotis palustris</i>	40
Figure 16	Degraded grasslands	41
Figure 17	A sketch of utilized flight lines	44
Figure 18	AISA DUAL sensor	45
Figure 19	AISA EAGLE sensor	46
Figure 20	AISA HAWK sensor	47
Figure 21	An illustration of the relations between different radiances (at a sensor or at surface)	48
Figure 22	Workflow for pre-processing of hyperspectral imagery in GEODIS	49
Figure 23	Spectra overlap of AISA EAGLE and AISA HAWK sensors	50
Figure 24	Errors in the aerial imagery	52
Figure 25	An example of a set of pixels prepared for the experiments	54
Figure 26	Mapping inaccuracies	56
Figure 27	An example of not enough detailed field mapping	57
Figure 28	The different amounts of pixels tested in the experiments	63
Figure 29	Random sampling design	65
Figure 30	Clustered sampling design	66
Figure 31	“Real life scenario” sampling design	67
Figure 32	Random pixels at boundaries	68
Figure 33	Classification maps for Přední Rennerovky for SVM polynomial kernel and SVM radial basis function kernel	82
Figure 34	Classification maps for Přední Rennerovky for SVM linear kernel and best combination of NN classifier parameters	83
Figure 35	Classification maps for Přední Rennerovky for randomly sampled 10% of total amount of pixels and clustered 5% of total amount of pixels	84
Figure 36	Classification maps for Přední Rennerovky for “real life scenario” 50 pixels and 10% pixels at boundaries	85

List of tables

Table 1	The most significant absorption features related to leaf components	13
Table 2	The area of the classified enclaves	32
Table 3	A comparison of AISA EAGLE and AISA HAWK sensors	46
Table 4	The examples of the studies comparing performance of SVM and NN classifier	58
Table 5	The summary of the kernels, the values of the penalty parameter and the amount of training pixels utilized in the studied literature	60
Table 6	The tested combinations of sampling design and amount of pixels	62
Table 7.A	The results of accuracy assessment for tested combinations of parameters of polynomial kernel	72
Table 7.B	The results of accuracy assessment for tested combinations of parameters of RBF kernel	72
Table 7.C	The results of accuracy assessment for linear kernel and different values of penalty parameter	73
Table 8	The results of accuracy assessment for tested combinations of sampling designs and the amounts of pixels	74
Table 9	The results of accuracy assessment for the location of Přední Rennerovky (SVM classifier)	75
Table 10	The results of accuracy assessment for the location of Friesovy Boudy (SVM classifier)	75
Table 11	The results of accuracy assessment for the location of Husí Boudy (SVM classifier)	75
Table 12	The results of accuracy assessment for the location of Klínové Boudy (SVM classifier)	75
Table 13	The results of accuracy assessment for the location of Lahrovy Boudy (SVM classifier)	76
Table 14	The results of accuracy assessment for the location of Zadní Rennerovky (SVM classifier)	76
Table 15	The results of accuracy assessment for tested combinations of parameters of the NN classifier	78
Table 16	The results of accuracy assessment for the location of Přední Rennerovky (NN classifier)	79
Table 17	The results of accuracy assessment for the location of Friesovy Boudy (NN classifier)	79
Table 18	The results of accuracy assessment for the location of Husí Boudy (NN classifier)	80
Table 19	The results of accuracy assessment for the location of Klínové Boudy (NN classifier)	80
Table 20	The results of accuracy assessment for the location of Lahrovy Boudy (NN classifier)	80
Table 21	The results of accuracy assessment for the location of Zadní Rennerovky (NN classifier)	81

1. Introduction

Mountainous ecosystems with plant communities are very sensitive indicator of global climate change and should be monitored. Moreover the variety of biotic and abiotic factors causes a diverse mosaic of habitats leading to high biodiversity and occurrence of endemites. That is the reason why such areas are often protected. Such precious areas should be also mapped often and thoroughly because the information can be used for their suitable management (Marcinkowska et al., 2014). Meadows in the Krkonoše mountains are, thanks to the specific type of their management, an example of such diverse ecosystem and they are one of the most precious ecosystems occurring in the Krkonoše mountains. Almost one third of all the species of original vegetation of the Krkonoše mountains can be found here, including endemites like *Campanula bohemica* (Pomahačová, 2012). However for maintaining such diverse biotope, it is necessary to continue with or re-establish regular mowing, grazing and other types of managements which help to maintain high biodiversity. In this matter it is also important to control if the planned management activity was realized or not and to map if it had any effect on the vegetation communities. Nevertheless due to low accessibility of terrain, very short vegetative season and unstable weather conditions mapping of the vegetation and its state are often constrained or even impossible using traditional, field mapping techniques (Jakomulská et al., 2003). Moreover visual mapping of vegetation in the field can be biased by subjectivity of a human factor. Therefore the techniques of remote sensing, which are not biased neither by human factor nor limited by terrain, are highly suitable for mapping of these valuable areas. In addition remote sensing measurements are repeatable under relatively same conditions (e.g. same area and same sensor) and in a comparison with the traditional field mapping also in a short time.

Meadow vegetation in the Krkonoše mountains is classified in this master thesis using aerial hyperspectral data and learning machine algorithms. The work is focussed mainly on the SVM classifier, but also NN classifier was utilized in the classifications. The main goals of the master thesis are to determine best setting of SVM parameters and to propose ideal design for a training dataset for the classification of the meadows in the Krkonoše mountains. Also the finding of Foody and Mathur (2004, 2006), that the SVM algorithm is able to classify datasets with the same or higher accuracy when border training pixels are used instead of the conventionally chosen ones (e.g. randomly sampled pixels) will be tested. The criterion of tests will be the result of classification accuracy. Results of these experiments

could be also used in the future as guidelines for field mapping, to limit extensive whole-area mapping to only “hot spots” (e.g. boundaries or centres of polygons). The additional goal of the master thesis is to compare performances of both utilized classifiers. Based on the studied literature, SVM classifier is anticipated to perform better than NN classifier and to achieve comparable accuracies of classification as NN classifier but with less training samples used.

2. Theoretical background

2.1 Remote sensing of vegetation

Vegetation is present on the majority of Earth's surface and together with physical environmental variables influences ecosystems and climate. Therefore vegetation is often subject of remote sensing. The reason, for which vegetation is so precious land cover type is photosynthesis. Photosynthesis takes place in chloroplasts, special lens-shapes structures in the palisade layer of leaves, where Chlorophyll molecules are located. Chlorophyll is the green pigment that is responsible for green colour of vegetation and also, absorbing sunlight, makes photosynthesis possible to happen (Campbell and Wynne, 2011).

Alongside pigments and other chemical substances present in leaves (nitrogen, cellulose, water, lignin etc.) also structure of leaves and their interaction with electromagnetic radiation influence spectral behaviour of vegetation (Jensen, 2007).

As plant usually has a bunch of leaves and plants are further grouped into assemblages, spectral properties of vegetation are also affected by interactions between leaves, stems and other components of these assemblages or canopies (Jones and Vaughan, 2010). Therefore it is important to distinguish between spectral behaviour of a leaf, a canopy or the whole community of plants.

2.1.1 Spectral properties of a leaf

The chemical composition of a leaf, especially photosynthetic pigments, determines its spectral properties in the visible part of spectrum (Jones and Vaughan, 2010). Regarding healthy leaves, chlorophylls are the most significant of all the photosynthetic pigments. However chlorophylls do not absorb all parts of visible spectra equally, their molecules preferentially absorb at blue and red wavelengths of spectrum (Campbell and Wynne, 2011; Jones and Vaughan, 2010). They may absorb as much as 70–90% of incident light in these regions (Campbell and Wynne, 2011). The carotenoids (lutein and β -carotene) and xanthophylls (violaxanthin and zeaxanthin) extend absorption of a leaf in blue – green region of visible spectrum. As leaves senesce in autumn or as a result of environmental stress, concentration of chlorophyll decreases and the photosynthetic pigments that absorb in blue – green part of spectrum (carotenoids, xanthophylls and anthocyanins) become more apparent. Therefore leaves change their colour in autumn or when subjected to environmental stress (Jones and Vaughan, 2010).

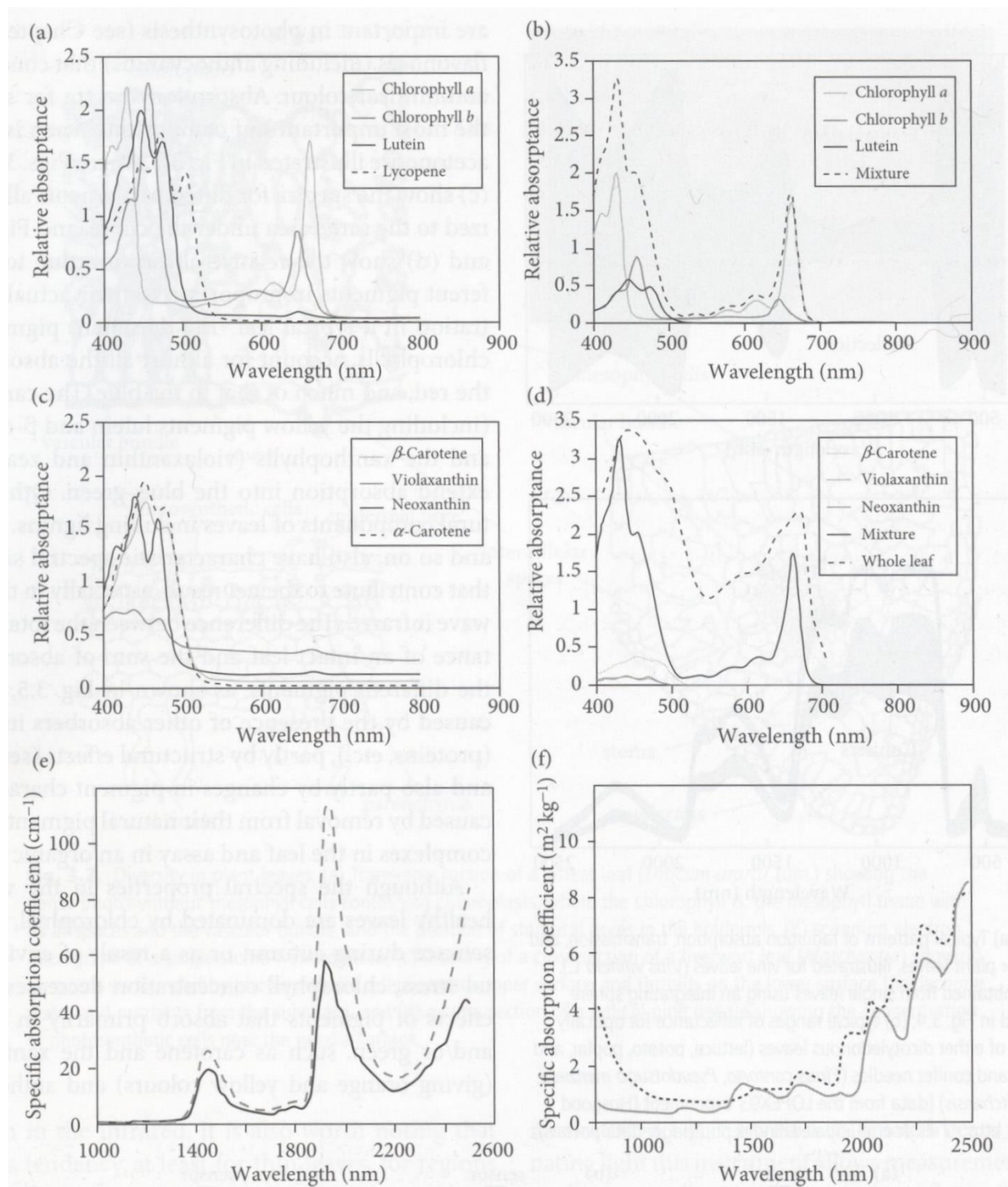


Fig. 1: Spectral features of key components of leaves. Absorption by chlorophyll-a, chlorophyll-b, lutein and lycopene (a) and β -carotene, Violaxanthin, Neoxanthin and α -carotene (c). Graphs b and d show absorption of different components scaled according to their typical concentrations in a leaf. Graph e shows comparison of the absorption of leaf water (solid line) and pure water (dashed line). Graph f shows absorption coefficients for leaf protein (dashed line) and cellulose together with lignin (solid line) (Jones and Vaughan, 2010).

The absorption of key pigments in visible spectrum to NIR is in fig.1, the specific absorption peaks of main absorption features are in the table 1.

Table 1. The most significant absorption features related to leaf components (Jones and Vaughan, 2010).

Chemical	Wavelength (nm)	Electronic transition or bond vibration
Chlorophyll	430, 460, 640, 660	Electronic transitions
Water	970, 1200, 1400, 1940	O-H bond stretching
Protein, nitrogen	1510, 2180	N-H stretching and bending, C-H stretching
Oil	2310	C-H stretching and bending
Lignin	1690	C-H stretching
Cellulose and sugar	1780	-

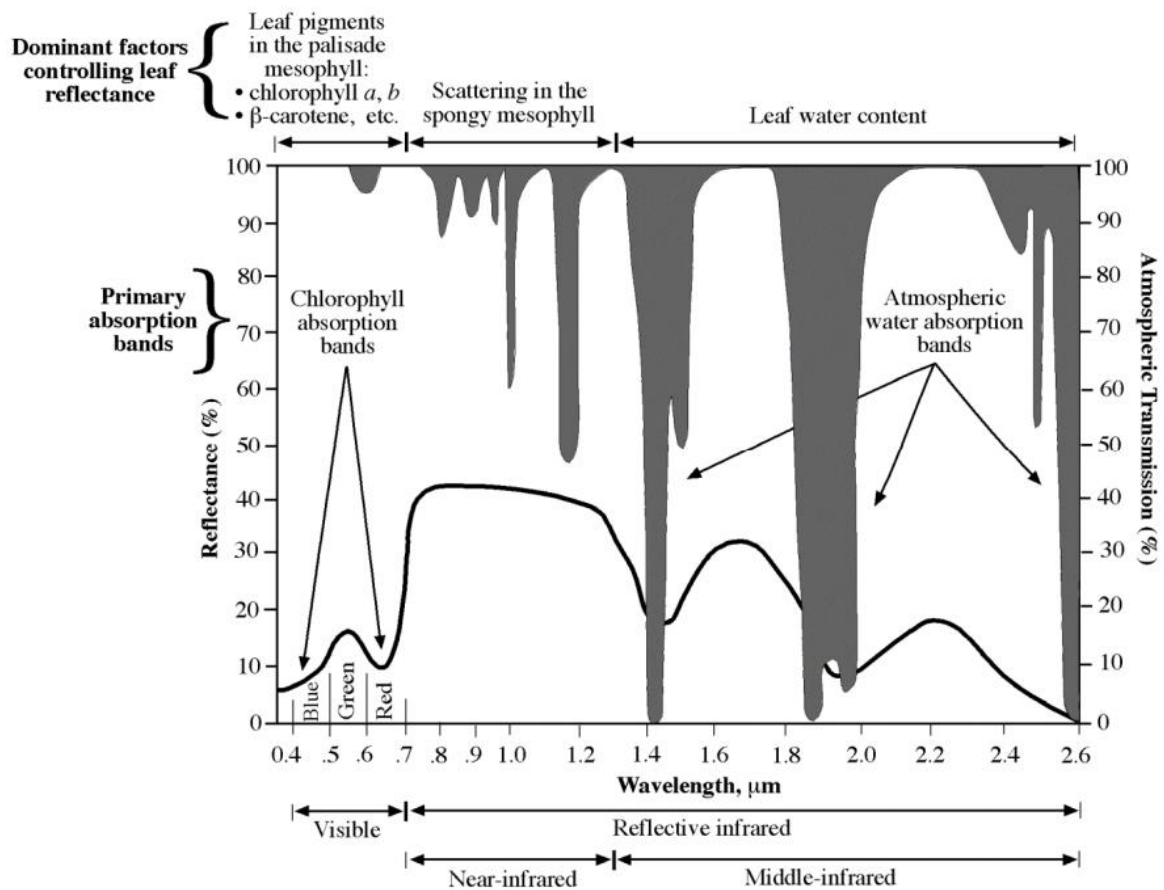


Fig. 2.: A spectral curve of a leaf (Jensen, 2007).

In the near infrared spectrum, reflection of the leaf is controlled neither by pigments nor by other chemical substances (Jones and Vaughan, 2010) but by the structure of the spongy mesophyll tissue (Campbell and Wynne, 2011) and other components in plant cell walls (Jones and Vaughan, 2010). As the cuticle and epidermis are almost completely transparent to IR radiation, only small portion of it is absorbed internally and the majority of it is either reflected or transmitted. Consequently NIR reflectance of healthy leaves is very high (Campbell and Wynne, 2011).

The sharp change in absorption / reflectance of a healthy leaf between visible and NIR (around 700 nm) is called red-edge and is the most important spectral feature in remote sensing of vegetation in general (Jones and Vaughan, 2010). Apart from separating vegetated areas from non-vegetated (Campbell and Wynne, 2011; Jones and Vaughan, 2010), this feature plays the key role also in discrimination between plant communities and single species (Campbell and Wynne, 2011). However its precise allocation on a spectral curve is possible only using hyperspectral data (Jensen, 2007).

In contrast to visible region of spectrum, water is the dominant chemical contributor to the radiative properties of vegetation in short-wave infrared (SWIR) (Campbell and Wynne, 2011; Jones and Vaughan, 2010). According to Jones and Vaughan, water absorbs strongly at wavelengths longer than about 1100 nm (Jones and Vaughan, 2010) but according to Campbell and Wynne the absorption occurs at even longer wavelength (about 1300 nm) (Campbell and Wynne, 2011). Jones and Vaughan further specify the absorption bands of water as: c.1450 nm, 1950 nm and 2500 nm (Jones and Vaughan, 2010), which is in agreement with Jensen 2007 (Fig.2).

2.1.2 Spectral behaviour of canopies – “hotspot” phenomenon

Regarding spectral behaviour of canopies, we cannot consider only interaction of incidence radiation with leaves but also with stems and other canopy elements (e.g. underlying soil) of canopies. Canopy reflection is also influenced by scattering of radiation and secondary and tertiary interactions between its components at different levels of the canopy. These depend on the detailed architecture or spatial organization in relation to the angular distribution of the incident radiation and the orientation of a sensor (Jones and Vaughan, 2010).

Consequently, the scaling up of information from single leaves and other canopy parts to derive the radiative properties of whole canopies is a complex task (Jones and Vaughan, 2010).

Important features of canopies are shaded areas (Campbell and Wynne, 2011; Jones and Vaughan, 2010). Vegetation canopies are composed of many separate leaves varying in their size, orientation, shape and coverage and leaves (and other components) are further assembled into many layers, of which the upper ones form shadows that mask the lower ones (Campbell and Wynne, 2011). Therefore the overall reflectance of a canopy is a combination of reflectance of its elements (leaves, stems etc.) and reflectance of shadows (Campbell and Wynne, 2011; Jones and Vaughan, 2010). Shadowing tends to decrease canopy reflectance below the values normally observed in the laboratory for individual leaves. As a result, the reflectance of a canopy is considerably lower than reflectance measured for individual leaves (Campbell and Wynne, 2011).

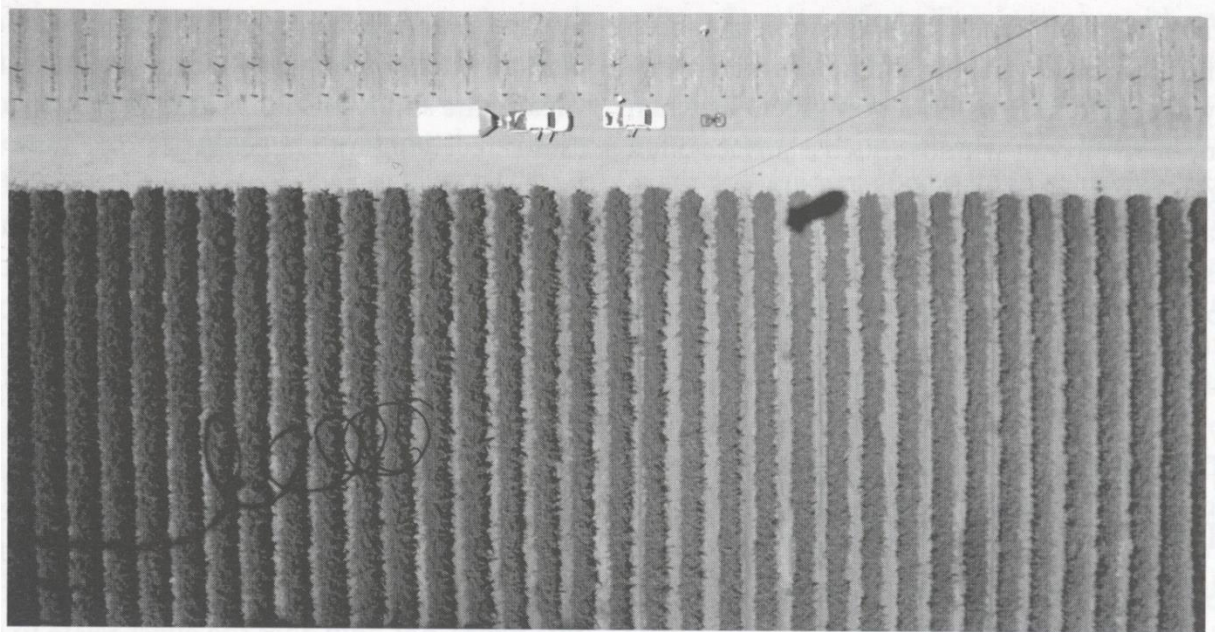


Fig.3: An aerial image of a vineyard from a balloon, showing the changing brightness of the image (Jones and Vaughan, 2010).

Moreover, shadows in canopies in connection with the changing of viewing and illumination angles cause also another phenomenon, called “hotspot phenomenon” (Fig.3). In the fig.3 we can clearly see, that vegetation nearer to the hotspot (the area around the shadow of the balloon) displays higher brightness than the vegetation further from the hotspot. The

reason for this is increasing ratio of shadows to vegetation visible, as one moves away from the hotspot (Jones and Vaughan, 2010).

The origin of bidirectional pattern of reflection (BR) from a plant canopy is illustrated in fig 4. Fig. 4 shows a situation when an array of plants is illuminated by direct sunlight. As plant canopies are complex three dimensional structures, some proportion of a canopy is in direct sunlight while some is in shadow. The proportions of sun and shade change with the view and illumination angle. When we look at a surface (canopy) with the sun behind us, majority of a surface is sunlit, but when we look at it while looking towards the sun much of the viewed surface is shadowed. This leads to before mentioned changes in the brightness of an image creating a “hotspot”.

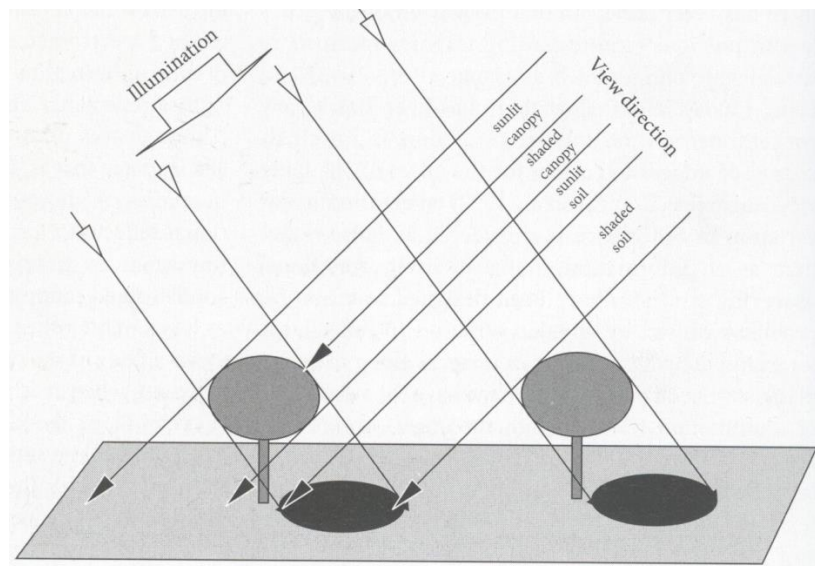


Fig.4: A hotspot diagram (Jones and Vaughan, 2010).

The reflectance properties of a surface are fully described by the bidirectional reflectance distribution function (BRDF). The BRDF is a mathematical description of the optical behaviour of a surface with to illumination zenith and azimuth angles and view zenith and azimuth angles (Campbell and Wynne, 2011; Jones and Vaughan, 2010). The BRDF is also commonly a function of wavelength (Jones and Vaughan, 2010).

2.2 Hyperspectral data in vegetation mapping

The term “hyper” has its origin in Greek language and means “over”, “above” or “exaggerated amount”. In contrast the term “multi” means “many”. These terms combined with “spectral”, which is related to the word “colours”, are combined to the words “hyperspectral” and “multispectral”. So the term “hyperspectral” can be translated as “exaggerated amount of colours” while the term “multispectral” means “many colours” (Borengasser et al., 2008). By this definition of the terms the first difference between multispectral and hyperspectral data is given – whereas multispectral imagery contains only units of bands, hyperspectral imagery usually contains hundreds of bands. In addition, hyperspectral bands are narrower than multispectral bands (units of nanometres versus tenths of nanometres) (Cambell and Wynne, 2011) and continuous (Cambell and Wynne, 2011; Borengasser et al., 2008). The amount, the narrowness and the continuousness of hyperspectral bands are the strengths and the weaknesses of hyperspectral imagery. On one hand, thanks to such detailed information even slight variations in spectral features of land cover can be noticed (Borengasser et al., 2008) and land cover types can be mapped with higher accuracy (Magiera et al., 2013), on the other hand there is lots of redundant information (as many of bands are often correlated) (Jones and Vaughan, 2010). The amount and the detail of information also mean higher demand on computer hardware and software and difficulties with storage and visualizations of such data (Jones and Vaughan, 2010; Borengasser et al., 2008).

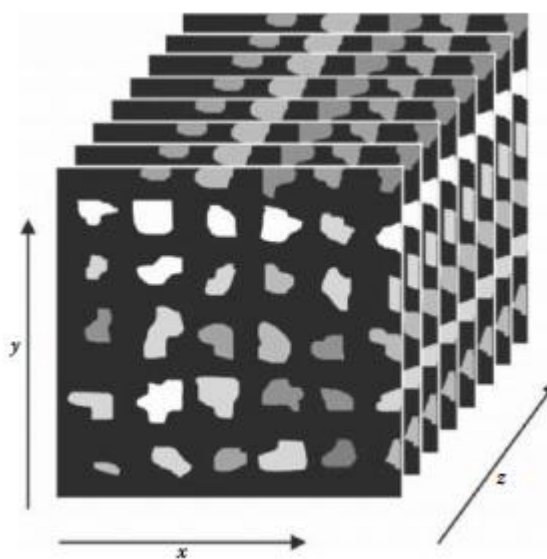


Fig.5: A hyperspectral cube (Borengasser et al., 2008).

Hyperspectral data is commonly visualized using a hyperspectral cube (or an image cube) (Campbell and Wynne, 2011; Borengasser et al., 2008) (Fig. 5). A spectral hyper-cube is a three-dimensional array containing spatial information of an image on the x and y axes and spectral information of an image on the z axis (Borengasser et al., 2008).

2.2.1 Methods for reducing the dimensionality of hyperspectral data

As it was mentioned before hyperspectral data contains many narrow bands, which are often correlated and carry lots of extra information. Additionally, classification methods are usually iterative (Magiera et al., 2013; Verrelst et al.), thus performing iterative methods with such mass of data would be very demanding for hardware. Therefore it is necessary to reduce the dimensionality of hyperspectral data (Borengasser et al. 2008, Feilhauer et al. 2010).

2.2.1.1 Principal component analysis (PCA)

The main purpose of PCA is to generate a new reduced and uncorrelated set of bands, while maintaining the same amount of information as in the original image. Usually the first principal component contains about 92% to 98% of the total information and typically over 99% of information is contained in the first three components (Jones and Vaughan, 2010). The last PC bands appear noisy because they contain very little variance, much of which is caused by noise in the original imagery (ENVI, PCA).

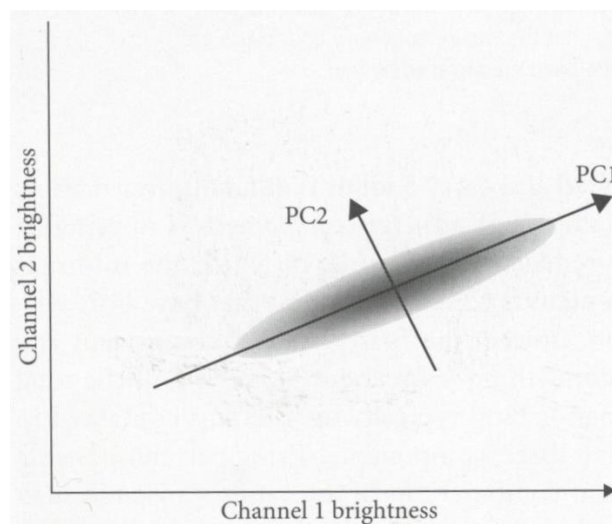


Fig.6: Schematics of the new dataset found by PCA (Jones and Vaughan, 2010).

Generating the new set is done by finding a group of orthogonal axes that have their origin at the data mean and that are rotated so the data variance is maximized (ENVI, PCA)

(Fig.6). These axes, or principal components are then linear combinations of the original bands, chosen in such a way that they are uncorrelated with each other (Jones and Vaughan, 2010).

2.2.1.2 Minimum Noise Fraction Transform (MNF)

The MNF algorithm, like PCA, is a linear, data reduction transformation. This transformation consists of two separate principal component transformations (ENVI, MNF).

In the first rotation the principal components of the noise covariance matrix are used to decorrelate and rescale the noise in data by variance (this process is also known as noise whitening). The transformed data has no band-to-band correlation (ENVI, MNF).

The second rotation is a standard principal component rotation of the noise-whitened data. Eigenvalues are then assigned to associated bands, which can be divided into two parts: one part with large eigenvalues and coherent images and a complementary part with near-unity eigenvalues and noise-dominated images (ENVI, MNF).

2.3 Classification methods currently used for grassland vegetation mapping

There are many classification methods available nowadays, but not all are fully suitable for classification of vegetation from several reasons (Magiera et al., 2013; Sha et al., 2009; Dobrowski et al., 2008; Schmidlein and Sassin, 2004). At first basic classification methods, which rely only on spectral information from an image without any ancillary data are likely to be less successful, especially when using multispectral imagery and/or when a classification is performed in mountainous areas. Because the same vegetation community may have different spectra in different parts of an image (e.g. due to various topography, dissimilar percentage cover of particular vegetation on a particular plot or due to BRDF caused by backscattering of vegetation), in contrast different vegetation communities may have similar spectra (Sha et al., 2009). This problem could be partially solved by using hyperspectral data, which has higher spectral resolution and thus it is possible to detect faint spectral differences of grassland vegetation types (Magiera et al., 2013). However, the problem of having different spectra for the same vegetation community due to topography or dissimilar percentage cover will not be solved by using hyperspectral data only. Some ancillary data is usually required to successfully map vegetation (Jones and Vaughan, 2010; Sha et al., 2009) and therefore also methods which are able to utilize such data, such as decision trees and ensemble classifiers (Jones and Vaughan, 2010; Lucas et al., 2007).

2.3.1 Probabilistic approach

Standard classification methods use „hard classifiers“, where each pixel is assumed to be “pure” (contains only single class) (Jones and Vaughan, 2010). However in vegetation communities (especially in such diverse communities as are meadows), there are hardly all pixels pure and lots of mixed pixels occur (Jones and Vaughan, 2010, Sha et al., 2009). These mixed pixels occur either at borders of biotopes or if another class is contained within a pixel – for example isolated trees in grasslands. Mixed pixels negatively influence the result of classification, where they are usually miss-classified into one of the surrounding classes or marked as unclassified. Therefore “soft classifier” methods, such as fuzzy classification may be better for classification of vegetation. Fuzzy classification utilizes probabilistic approach to classify a pixel and allows “partial membership” of classes per pixel. The output of fuzzy classification is set of fractional images, which describe the probability of presence of each vegetation community within each pixel (Jones and Vaughan, 2010). The logical decision

would be, that the class with the highest probability in particular pixel is contained in this pixel, however Sha et al proved in their study that not every time the class with the highest probability of presence in particular pixel actually was in the pixel and in several cases the classes with the second highest probability occurred in the pixel (Sha et al, 2009). Therefore in the future it would be convenient to take into account also classes with the second highest probability of presence. In case that classification of image using highest probability fractional images fails to classify vegetation communities correctly.

Another probabilistic approach to classification of mixed pixels of vegetation is linear spectral unmixing (LSU). This method is based on the idea that each pixel contains linear combination of reflectance spectra of all end-members (Borengasser et al., 2008). Furthermore the maximum amount of end-members (component classes) identifiable in a pixel is limited by the number of spectral bands available (Jones and Vaughan, 2010). Therefore it is clear that linear spectral unmixing is used in connection with hyperspectral data (because of abundant spectral bands). The multispectral data would not have probably enough spectral bands to successfully identify mixed pixels of grasslands. Another minor disadvantage of this method is the presumption that end members in a pixel are linearly combined, which, in my opinion might not be true in many cases.

2.3.2 Gradient modelling

Schmidtlein and Sassin argue in their work, that fuzzy classifiers and linear spectral unmixing methods are suitable for classifying transition zones of vegetation communities if these transitions represent a mixture of ideal types of species assemblages (end-members). But according to Schmidtlein and Sassin these ideal types does not always have to occur in scene even though their existence is assumed and this idea is not in agreement with case of continua, where any section of a gradient is considered to be pure (Schmidtlein and Sassin, 2004). Concept of continua is ecological approach to classify vegetation, based on the theory that transitions between vegetation units are always continuous and based on the slope of environmental gradients (Verrelst et al., 2009, Schmidtlein and Sassin, 2004). “The clash” between ecological (botanical) and remote sensing approach in vegetation mapping is obvious at this point. Whereas remote sensing usually divides vegetation into groups based on their spectral discernibility, ecologists and botanists divide vegetation into either ecologically or botanically meaningful groups, which, however does not have to be spectrally distinguishable

(Magiera et al., 2013; Verrelst et al., 2009, Schmidlein and Sassin, 2004). Therefore methods are being tested currently how to assemble plants into communities in a way that these communities maintain ecological or botanical meaning while they are possible to differentiate by remote sensing data (Schmidlein and Sassin, 2004). Organizing plant community data along environmental gradients is also called as species distribution modelling and usually requires an extensive amount of predictor variables, and therefore use of multivariate statistical methods such as ordination are necessary (Dobrowski et al., 2008). An Ordination analysis seeks to detect sets of variables that accounts for the major patterns across all the original predictor variables without a substantial loss of information (Verrelst et al. 2009). The predictor or environmental variables are such attributes which determine species distribution. These are not only physical or chemical environment variables (i.e. elevation, slope, aspect, temperature, bedrock geology, soil reaction or occurrence of pollutants) but also ecological and biophysical variables (i.e. competition or symbiosis of species; LAI and chlorophyll content). Obviously ancillary data (i.e. digital elevation model, geological maps, result of chemical analysis of soils) are necessary for deriving environmental physical variables and some level of ecological knowledge is necessary to determine relationships between species. When employing biophysical variables as predictor variables, it is necessary to measure them in-situ by using specialized technology or they can be derived from vegetation indices. Even though the gradient modelling can predict well species distribution, it characterizes more potential than an actual vegetation distribution and therefore it is used in combination with classical image analysis (Dobrowski et al. 2008).

2.3.3 Vegetation indices

Vegetation indices have been also used for classifying vegetation. Vegetation indices are new variables created by mathematical combinations of 2 or more of the original spectral bands of an image. These variables are created in the way, that they are better related to biophysical parameters of interest, than when only spectral bands are used (Roberts et al., 2011; Jones and Vaughan, 2010). Biophysical parameters assessed by vegetation indices can be divided into 3 major groups – structure (i.e. fractional cover, green leaf biomass, leaf area index or fraction of absorbed photosynthetically active radiation); biochemistry (i.e. water, pigments, nitrogen-rich compounds or cellulose and lignin content) and plant physiology (i.e. changes in chlorophyll content or water moisture in leaves) (Roberts et al., 2011). The indices

can be divided into simple indices (such as Difference vegetation index or Ratio vegetation index) and into normalized indices (i.e. Normalized difference vegetation index or soil adjusted vegetation index). Simple indices are more sensitive to changes in illumination conditions and also perform worse when lower vegetation percentage cover occurs (lower LAI). Therefore normalized indices were created in order to better manage these problems. For example SAVI (soil adjusted vegetation index) is more closely related to LAI over a range of soil reflectance than normalized vegetation index (NDVI) (Jones and Vaughan, 2010). On the other hand, when chlorophyll content is high (and thus higher LAI) the saturation of NDVI occurs, because the range of chlorophyll content is higher than what NDVI is able to cope with (Wang et al., 2007). Therefore it is possible to use for example Green NDVI, where red bands are exchanged for green bands and thus saturation does not occur and estimation of LAI gives better results (Jones and Vaughan, 2010; Wang et al., 2007). For improving accuracy of prediction of biophysical variables and better distinguishing particular vegetation species or communities, hyperspectral data has been lately used to construct more accurate vegetation indices (Roberts et al., 2008; Jones and Vaughan, 2010).

2.3.4 Machine learning algorithms

Last but not least, “machine learning” classification algorithms, such as artificial neural networks (or neural networks; ANN) or support vector machines (SVM), are also part of “pixel based” classifiers (Petropoulos et al., 2012, Camps-Valls et al., 2004). Both methods are supervised non-parametric methods, which means that there is no assumption made about the data distribution (e.g. they do not presume normal distribution of data), which is on the contrary of other conventional classifiers, such as maximum likelihood classifier (Jones and Vaughan, 2010). This fact is a big advantage of ANN and SVM because remotely sensed data have, in majority, unknown statistical distribution. Furthermore hyperspectral remote sensing data have usually distribution far from normal because as dimensionality of data increase, its density tends to be higher in tails (Mountrakis et al., 2011, Melgani and Bruzzone, 2004).

2.3.4.1 Neural Networks

ANN is algorithm designed to simulate human learning process by establishing linkages between input and output data via one or more hidden layers (fig.7). The algorithm is

reinforced by repeated learning (Jones and Vaughan, 2010). Basic unit of each layer in ANN set-up is neuron (node). The neuron has many inputs, while only one output. Input signals represent an activity of neural impulses delivered by another neuron to this particular neuron (Benediktsson et al., 1990). Neurons are linked by weighted connections according to a specified architecture. Whether the signal is passed from particular neuron to another neuron depends on a magnitude of input signals and on an activation function. If a weighted sum of all incoming signals according to activation function exceeds given threshold, the neuron is activated and the signal passes to a following neuron (Petropoulos et al., 2012; Benediktsson et al., 1999). Learning in ANN algorithm occurs by adjusting the weights in the nodes and optionally the node thresholds to minimize the difference between the output node layer and desired response (Petropoulos et al., 2012).

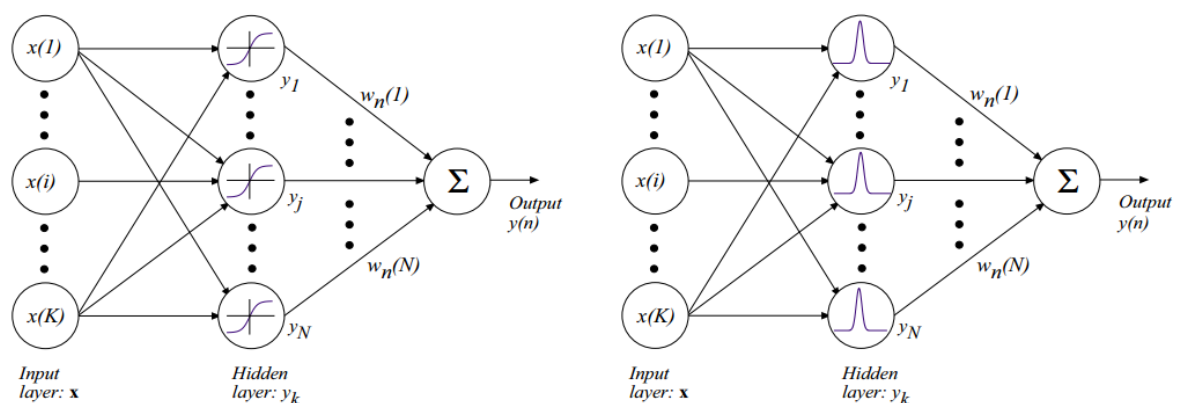


Fig. 7: A schema of a neural networks setup. MLP setup on the left (with sigmoidal activation function) and RBF setup on the right (with Gaussian activation function) (Camps-Valls et al., 2004).

Regarding the set-up of ANN network, there is always one input layer, one output layer and one or more hidden layers. The nodes in the input layer represent variables used as an input into neural network, such as spectral bands, textural features or other features derived from a remotely sensed image. The nodes of the output layer represent classes, there is one node for each class (Petropoulos et al., 2012). The hidden layer(s) are made up of multiple nodes, which number depends solely on a network architecture and complexity of classification task and differs greatly throughout studies from 26 (Pal and Mather, 2005) to 93 (Chen et al., 2012). There are several models of ANN based on how neurons are connected or which activation function is used. The classic model of a feed-forward multilayer neural network, known as multilayer perceptron (MLP) has fully-connected neurons between all

layers (input, output and hidden), which means that each neuron of given layer feeds all the neurons in the next layer (Camps-Valls et al., 2004). Most common activation function in MLP is sigmoidal or hyperbolic activation function (Petropoulos et al., 2012). And this model is also the most commonly used in classifications (Huang et al., 2002). In radial basis functions neural networks (RBF) layers are connected in the same manner as in previously mentioned model, however the sigmoidal or hyperbolic function is replaced by Gaussian function. According to Camps-Valls et al., this model tends to overfit when it has to deal with noisy inputs (Camps-Valls et al., 2004). Lately also neuro-fuzzy approach has been applied, in which neural networks principles are combined with fuzzy logic. The components of such NN are layer of neurons, which apply fuzzy membership function to inputs, and modular network, which applies functional rules to the inputs. In the end the combiner needs to be used to apply membership function outputs to modular network outputs. The best classification results are according to Camp-Valls et al., given by neuro-fuzzy models, especially when classifying vegetation. Fuzzy component helps to catch soft boundaries between particular vegetation classes; however these are also the most complicated and the most difficult to set-up (Camps-Valls et al., 2004). According to several authors, the model with one hidden layer is sufficient for the majority of classification tasks (Petropoulos et al., 2012; Camps-Valls et al., 2004; Pal and Mather, 2005; Huang et al., 2002; Benediktsson et al., 1999).

2.3.4.2 Support Vector Machines

Another machine learning algorithm is support vector machines algorithm (SVM) (fig.8). This algorithm is based on use of statistical learning theory and aims to find the best hyperplane in a multidimensional feature space that optimally separates classes (Jones and Vaughan, 2010; Mountrakis et al., 2011). The term best hyperplane is used to refer to a decision boundary obtained in a training step and minimizing misclassifications (Petropoulos et al., 2012; Mountrakis et al., 2011). According to Chan et al. good classifier should be capable of generalization to correctly assign classes based only on few training samples (Chan et al., 2012). According to Camps-Valls et al and Mountrakis et al., the SVM algorithm has good generalization capabilities because the separation hyperplane is constructed based on only fragment of all training examples (Mountrakis et al., 2011; Camps-Valls et al., 2004). Training examples used for construction of hyperplane are called support vectors. Support

vectors lie on the margin of classes to be classified and are extracted automatically by algorithm (Petropoulos et al., 2012; Mountrakis et al., 2011, Camps-Valls et al., 2004). According to Petropoulos et al. using only fragment of all the training samples to construct hyperplane is also measure against computational overload of algorithm (Petropoulos et al., 2012).

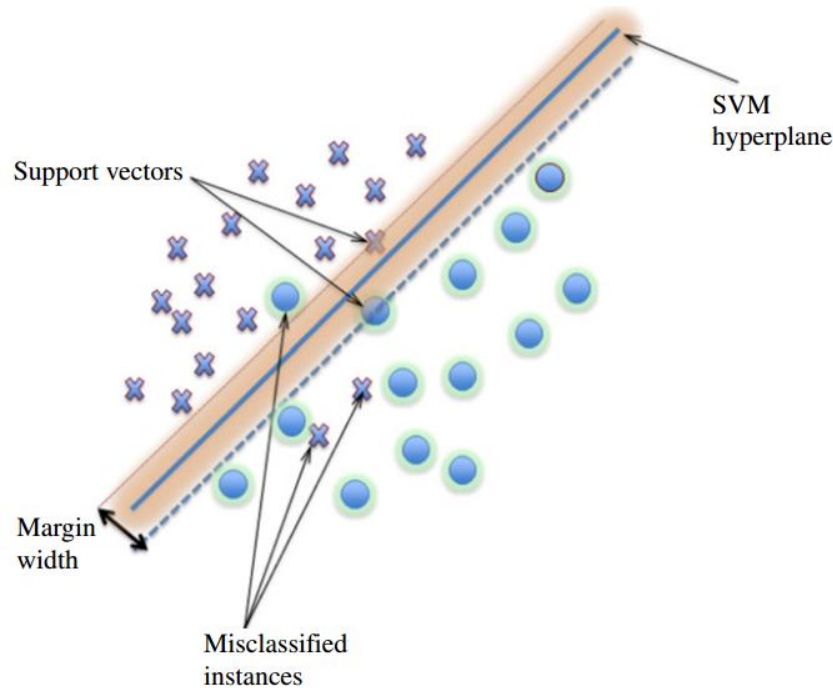


Fig.8: An example of linear SVM (Mountrakis et al., 2011).

The basic type of SVM – linear SVM assumes linear separability of classes. However this is not usually a case of real life scenarios, where clusters of different classes overlap one another (Mountrakis et al., 2011). This problem is in SVM algorithm solved by two ways, which can be applied separately, but mostly they are applied both. First is introducing so-called slack variables, in other words allowed errors. The magnitude of allowed errors is determined by parameter C , chosen by user, in the way that larger C corresponds to assigning higher penalty to errors (Camps-Valls et al., 2004). This method is as well called soft-margin method (Mountrakis et al., 2011). The parameter C is not positively given yet and in assessed studies has always been found by experiment (e.g. Petropoulos et al., 2012, Vyas et al., 2011, Camps-Valls et al., 2004). Another way of mapping non-linear relationships between clusters of different class memberships is using of kernels. By using kernel the input data are transformed into higher dimensional space where they are linearly separable (Camps-Valls et al., 2004) (fig.9). There are several types of kernels, which can be used for this task, such as

linear, polynomial, sigmoid or Gaussian (radial basis function). A kernel and its parameters have to be chosen carefully as kernel function and kernel parameters often have a bearing on the results of analysis. Too low values may lead to over-fitting, while too high values may lead to over-smoothing (Mountrakis et al., 2011).

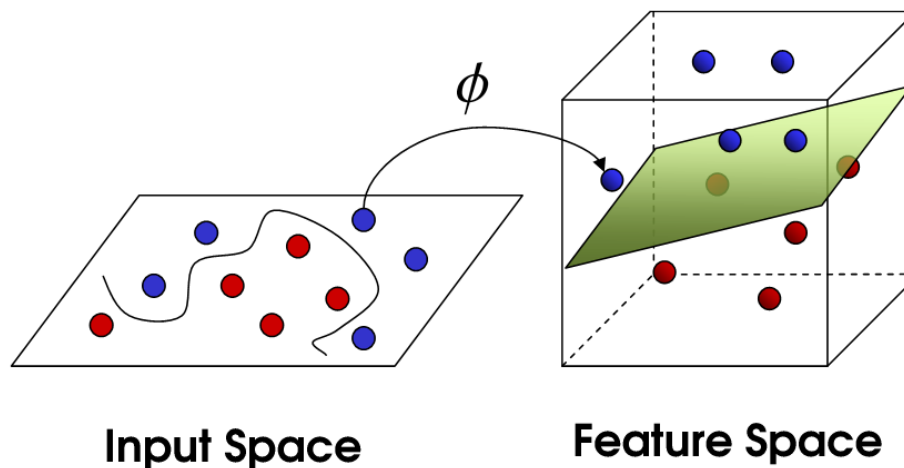


Fig.9: The transformation of input space into higher dimensional feature space and constructing a separating hyperplane. The variable ϕ symbolizes non-linear kernel function (Redd.it, machine learning).

SVM algorithm was originally constructed as binary algorithm. However, the classification of land cover usually involves a simultaneous discrimination of numerous classes (Melgani and Bruzzone, 2004). Therefore there have been developed principles how to extend this classifier also for classification of more than 2 classes –a one-against-one strategy and one-against-all strategy (Petropoulos et al., 2012, Melgani and Bruzzone, 2004). In the one-against-one strategy, $N*(N-1)/2$ SVM classifiers are created (N is number of information classes), in this way all possible pair-wise classifications are modelled. These classifiers are then processed in parallel to get a final result (Melgani and Bruzzone, 2004). In one-against-all strategy, one particular class is trained against all other classes and this is repeated for all the N information classes (Petropoulos et al. 2012). The classifiers are then again processed in parallel to final result. According to Melgani and Bruzzone the problem of one-against-all strategy is that discrimination between an information class and all the others often leads to the estimation of complex discriminant functions. Furthermore Melgani and Bruzzone investigated the influence of multi-class strategies on the accuracy of classification and also on computational time. Both strategies yielded comparable results regarding the classification, but computation time of the one-against-all strategy was higher by one order

(hundred seconds for the one-against-one strategy versus thousands seconds for the one-against-all strategy) (Melgani and Bruzzone, 2004).

To sum up, both ANN and SVM are supervised machine learning classifiers, the main strength of which is that they do not make presumptions about statistical distribution of data classified. SVM use only a fragment of training data, namely support vectors, which are samples defining margins of classes and thus separating hyperplanes. Therefore there is no requirement to fully describe the classes in a feature space, but only their decision boundaries. This is different from other conventional classifiers which usually require a full description of classes (Foody and Mathur, 2004). Foody and Mathur furthermore proved that for quite accurate classification only 35 of actual 150 training samples were necessary. Thus only a small training set is needed for SVM to perform satisfactory. This is welcome advantage because gathering training and also testing samples is always tedious and very time consuming work (Foody and Mathur, 2004). On the contrary ANN needs more training samples for successful training than SVM (Benediktsoon et al., 1990). Additionally there is no need for repetition of classifier training, such as for ANN (back-propagation of an error). Even though SVM seems as better performing classifier than ANN in several ways, there is one major drawback – the choice of kernel (Mountrakis et al., 2011). As it was explained before there are many types of kernels and choice of the kernel has impact on the result of analysis. But there are no guidelines so far, for using specific kernels and this matter was not yet discussed much in literature, therefore this matter still remains a question (Belousov et al., 2002).

2.3.5 Object-oriented classification

The before mentioned methods are all “pixel based” and do not take into account contextual or topological information (Debeir et al., 2002). Therefore an object-oriented image classification (OBIA) was developed. According to Dobrowski et al., it is possible to use this method when the spatial resolution of an image becomes finer and image structure becomes correlated to vegetation structures (Dobrowski et al., 2008). Homogeneous areas in an image are identified based on pre-defined attributes (spectral, textural or topological features) and these areas are then assigned to a particular class. This method is usable for example when end-members are not clear or available or are overlapping in a spectral space

(Jones and Vaughan, 2010). From the stated facts about the object-oriented classification it is obvious that this is one of the promising methods, which can be used for classifying grassland communities where spectral responses of some groups may be similar, while texture not or where some kind of topological relationship occurs (i.e. some vegetation communities are exclusively bind to others or on the other hand would never exist next to some). Nevertheless object-oriented classification is a “hard classifier” (Jones and Vaughan, 2010), therefore the continuous transitions between vegetation communities may be classified wrongly. Given this fact, it is important to decide whether it is more important to classify image with emphasis to continuous transitions between plant communities or if it is more profitable to use object-oriented techniques. And also as it was mentioned before it is possible to combine object-oriented classification with fuzzy approach or gradient modelling to achieve desired result.

2.4 Methods used for the vegetation classification in the Krkonoše Mountains

The meadows vegetation in the Krkonoše Mountains have been classified already as part of previous master and bachelor theses. The results of those are noted below, ordered chronologically.

Michaela Pomahačová (2012) classified meadows vegetation in the Krkonoše Mountains using multispectral data from satellite sensors WorldView 2 and Quickbird. A satellite sensor Quickbird is the traditional multispectral VNIR sensor, while sensor WorldView 2 has several bands added in comparison with Quickbird – coastal blue, yellow, red-edge and NIR 2. The additional bands were added to WorldView 2 sensor to improve different mapping analysis, vegetation included. Both of the sensors have very good spatial resolution (regarding other satellite sensors) – 1.84 m for multispectral data and 0.5 m for panchromatic data. The grasslands in Pomahačová’s thesis were classified according to 2 legends. The first legend was created by a botanist to encompass valuable or abundant species, while the second legend was inspired by NATURA 2000. The legend created by the botanist had approximately a half of the items similar to the legend created for my master thesis (stands dominated by *Vaccinium* species, waterlogged grasslands and degraded grasslands dominated by *N. stricta*), however the legend for Pomahačová’s thesis was created with the higher emphasis to grasslands under some management. The main method used for classification of grasslands was ANN and the conventional classifier MLC was used for the

comparison. There were always two images classified for each satellite sensor, the first from August 2011 and the second from September 2011. The overall accuracies achieved for Quickbird for August and September were 74.7% and 76.5% respectively. The overall accuracies achieved for WorldView 2 for August and September were 90.7% and 91.1% respectively. The accuracies achieved by conventional MLC classifier were lower in all the cases and therefore not noted here (Pomahačová, 2012).

Another master thesis, which was focused on the classification of meadows in the Krkonoše mountains was the one of Jan Jelének (2013). Jelének was using hyperspectral data from APEX sensor for the classification (Jelének, 2013). The sensor APEX is airborne pushbroom imaging spectrometer and is a result of a joint Swiss/Belgian project funded by ESA (Meuleman, 2012). APEX is able of recording hyperspectral data in approximately 300 bands in the wavelength range between 400 nm and 2500 nm. The spatial ground resolution of APEX is from 2 m to 5 m (APEX – ESA). The Jelének's thesis was focused on mapping invasive species, rather than on management or structure of vegetation and thus only one item classified was similar (grasslands dominated by *N.stricta*). The method used for classification was a decision tree (DT) with the use of vegetation indices (VI), which Jelének created himself based on the studied literature. Spectral angle mapper (SAM) was used as a reference method (Jelének, 2013). Based on the studied literature, this method was not used as frequently as other methods and also as it is shown further, this method was not successful in the mapping of mountainous grasslands, therefore it is not described in the chapter 2.3. The highest accuracy achieved for Jelének's classification by DT with use of VI was 67.9%, while with SAM only 21.4% (Jelének, 2013).

Roman Dorič (2013) was also mapping meadows in the Krkonoše Mountains. His work could be considered as a follow-up to Pomahačová's thesis because, besides the data from the optical airborne sensor, the data from the WorldView 2 sensor (imagery from September 2011) was used in the Dorič's thesis too. Nevertheless the legend was slightly different than in the Pomahačová's thesis, more similar to the legend created for my master thesis (approximately 80% of the items is similar). Dorič used object oriented classification with the help of SVM algorithm as a classifying method. The overall accuracy achieved was 59.0% for the WorldView 2 data and 54.1% for the airborne sensor data (Dorič, 2013). The work of Dorič could be assumed as the main reference work for my master thesis for two reasons, firstly – because of the similar legend and secondly – the SVM algorithm was used

as well. However there is one significant difference. Unlike in the Dorič's thesis, where solely multispectral data was utilized, hyperspectral data was used in my master thesis.

Martina Andrštová (2014) classified the biotopes quite similar to the meadows in the Krkonoše mountains – the non-forested areas above the treeline. However, as the biotope is not influenced by management as much as the meadows, the legend used in the Andrštová's thesis was different. Andrštová utilized several classification methods – SAM, LSU, SVM, MESMA (Multiple Endmember Spectral Mixture Analysis) and NN, and hyperspectral data from two sensors AISA and APEX. The accuracies achieved by SAM, LSU, SVM, MESMA and NN were 99.6%, 99.8%, 82.0%, 86.3% and 96.2% respectively. It was also confirmed that the data with the higher spatial resolution gives better results (Andrštová, 2014).

The non-forested areas above the treeline were also mapped in the bachelor thesis of Lucie Jakešová (2014). The legend used in this bachelor thesis was alike to the legend in Andrštová's thesis. Jakešová utilized data from an airborne optical sensor with 4 bands (blue, green, red and NIR band) and 12.5 cm ground spatial resolution. She also made use of DEM and DSM, which were derived from laser scanned data. The object oriented classification algorithm and vegetation indices were employed in this bachelor thesis, yielding the overall classification accuracy 76% (Jakešová, 2014).

Matouš Karvánek (2014) analysed land cover change and landscape state in the Krkonoše mountains between years 1999 and 2007, using ML, NDVI and landscape metrics calculations, thus his legend was broader than the legends utilized in the other mentioned theses. It contained traditional landscape categories, e.g. water areas, deciduous trees, coniferous trees and also a few additional vegetation categories such as stands with *Pinus mugo*, sparse woodland vegetation or grasslands. The analysis was performed with the use of archived airborne imagery of map server Mapy.cz. The accuracy of classification was on average 82% (Karvánek, 2014).

3. The study area

The study area is situated in The Krkonoše Mountains, in northeast part of Bohemia. Thanks to a mountainous landscape, the flora on the locations is diverse and range of individual communities varies greatly over the study area. The vegetation of The Giant Mountains belongs to a group of central European woodland flora (*Hercynicum*), with subgroup of *Sudeticum*. Furthermore, the vegetation of this group is usually not as rich as in The Krkonoše Mountains. The reason for this exception is the unique biogeographic position of the mountain range. Consequently, Krkonoše became a “home” for northern and alpine plant species during glacial periods, and thus great amount of glacial relicts can be found in this mountain range. As the area is quite isolated from other ones with similar climate a lot of endemic species developed there too. Therefore the flora of The Krkonoše Mountains is the most diverse of all the Central European mountain ranges (KRNAP, flora).

According to a preliminary plan, about thirty locations should have been classified. However, because of problems with field data, only six locations have been chosen in the end. The names of the study enclaves are as follows (fig.10): Friesovy Boudy, Husí Boudy, Klínové Boudy, Lahrovy Boudy, Přední Rennerovky and Zadní Rennerovky. The size of the classified area in total was 2 square kilometres. The partial areas of locations are in table 2.

Table 2: The area of the classified enclaves (source: own ArcGIS analysis).

Name of location	Area (km ²)
Friesovy Boudy	0.20
Husí Boudy	0.17
Klínové Boudy	0.51
Lahrovy Boudy	0.29
Přední Rennerovky	0.16
Zadní Rennerovky	0.60

3.1 Climate

The Giant Mountains form a natural barrier to moist westerly from The Atlantic Ocean. As a result, the climate of The Giant Mountains is more humid, colder and generally harsher than the climate of The Tatry Mountains or of The Šumava Mountains, moreover its character is rather oceanic. One feature of such climate is strong variability in very short time intervals (KRNAP, climate). What’s more, westward winds usually bring clouds and heavy

precipitation. Thanks to it there are only 30 to 40 sunny days and about 1500 sunlight hours throughout the year (KRNAP, sunlight). The strongest winds occur in winter and their speed can reach over 150 km/h (KRNAP, wind).

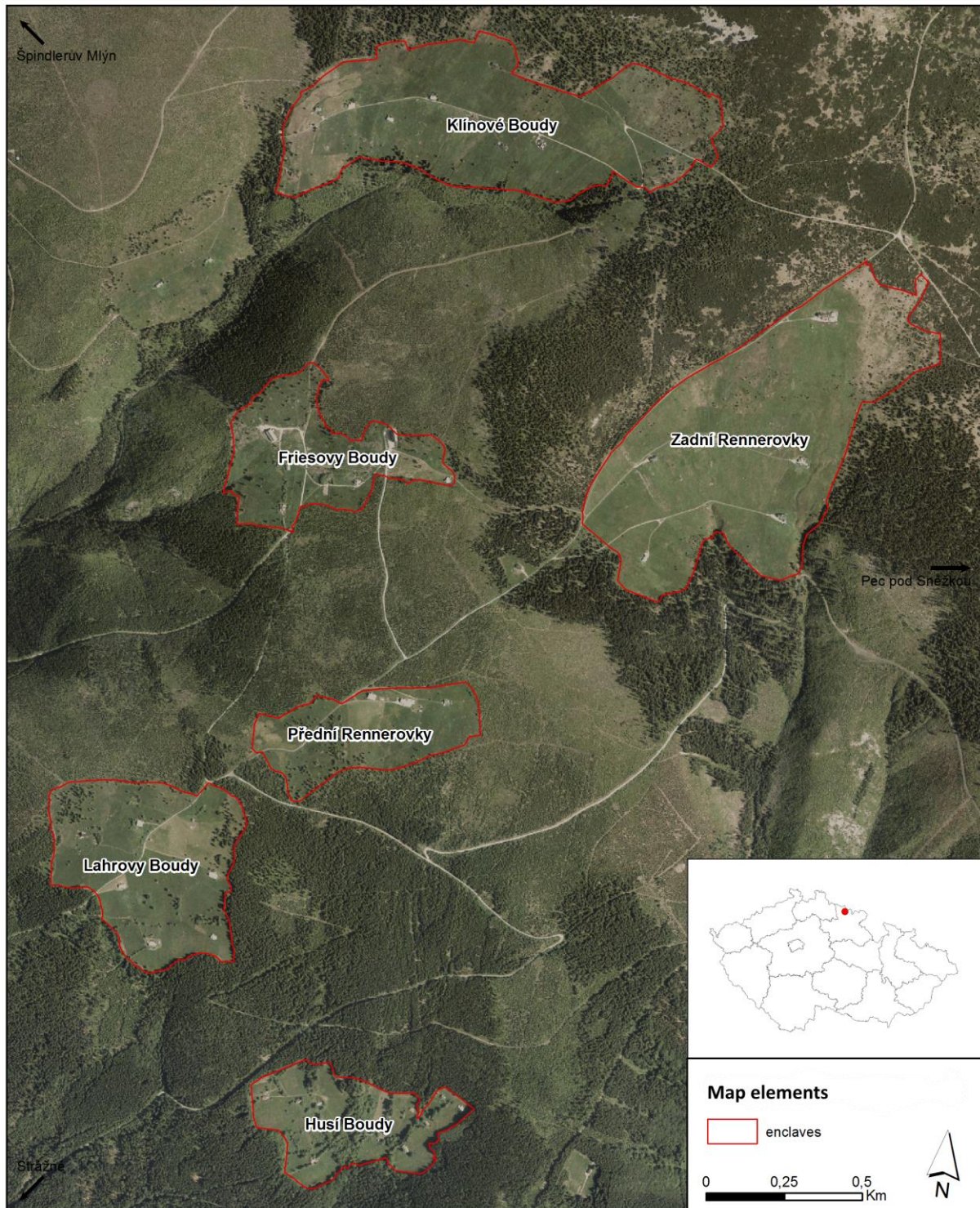


Fig. 10: A map of the classified enclaves (source: own ArcGIS output).

The average annual air temperature is between 6°C and 0°C (depending on a location). The warmest month is July (11°C on average) the coldest is January (-6°C on average). “Temperature inversions” are also regular in The Giant Mountains, especially during autumn and winter, lasting typically several days, rarely several weeks (KRNAP, temperature).

Precipitation increases with the increasing altitude. Snow, hail and hoarfrost are at high altitudes more common than rain. The average annual precipitation ranges from 800 mm to 1400 mm, with the highest frequency in August and the lowest in March (KRNAP, precipitation). Continuous snow cover is formed in November and its average depth is around 200 cm. Snow is accumulated on leeward slopes in winter and later causes snow avalanches during thawing periods. While avalanches are threat to tourists, they are important natural factor for biodiversity of the vegetation in The Giant Mountains (KRNAP, snow).

3.2 Classified types of biotopes

3.2.1 Legend

There are many ways how to categorize vegetation of the Krkonoše grasslands into classes, for example with regard to invasive plants, unique communities or management. However after consultations with the botanist of KRNAP, it was decided to create legend similar to some of the legends according to which grasslands in The Krkonoše Mountains have been already classified (e.g. in the master theses of Michaela Pomahačová or Roman Dorič) so the results of the classification are better comparable. The original legend was thus created by the botanist of KRNAP as follows:

1. Water surfaces; watercourses
2. Walls; piles of stones, rocks
3. Paved areas (objects, communications), dwellings, shacks
4. Solitary conifers and their self-seeded species
5. Solitary deciduous trees and their self-seeded species; scrubs
6. Stands dominated by *Vaccinium* species
7. Herbaceous vegetation under some management:
 - a. Stands dominated by *Nardus stricta*
 - b. Oligotrophic grasslands without dominance of *Nardus stricta*

- c. Mesophile grasslands
- d. Waterlogged grasslands
- 8. Herbaceous vegetation under none management:
 - a. Degraded meadows dominated by grass species (*Poales* except *Nardus species*)
 - b. Degraded meadows dominated by *Dicotyledons*

Furthermore, the first five items of the legend were masked to fully use the potential of SVM classifier for meadows classification (chapter 6.1.2.4) and based on several initial unsuccessful classifications, it was necessary to ignore categorization according to management, as multi-temporal data and field mapping during the time of acquisition would be needed to classify the meadows according to this aspect properly (Sulzer et al., 2013; Lucas et al. 2007). The resultant legend was then:

1. Stands dominated by *Nardus stricta*
2. Oligotrophic grasslands without dominance of *Nardus stricta*
3. Mesophile grasslands
4. Waterlogged grasslands
5. Degraded meadows dominated by *Dicotyledons*
6. Degraded meadows dominated by grass species (*Poales* except *Nardus stricta*)
7. Stands dominated by *Vaccinium* species
8. Springs
9. Subalpine tall-fern vegetation

This legend is very similar to the legend used in the master thesis of Roman Dorič (Roman Dorič, 2013), except classes 8 and 9, which have not been classified in Dorič's thesis.

3.2.2 Description of the classified biotopes

3.2.2.1 Closed alpine grasslands (with dominating *Nardus stricta*)

This biotope is comprised of heavily closed grasslands of mostly only one dominating species – *Nardus stricta* (fig.11). Closed alpine grasslands occur usually on flat mountains ridges, plateaus and on mild mountainous slopes, on oligotrophic but well-developed soils.

There are many complementary species (e.g. *Anthoxanthum alpinum*, *Carex bigelowii*, *Homogyne alpina*, *Luzula sudetica*, *Molinia caerulea* or *Potentilla erecta*) however these occur in minority. The biggest threat to this unique biotope is trampling by tourists and eutrophication (Chytrý et al., 2001).



Fig. 11: Closed alpine grasslands dominated by *Nardus stricta* (source: author).

3.2.2.2 *Nardus* grasslands (oligotrophic grasslands)

Nardus grasslands are oligotrophic grasslands with maximum height of 40 cm. We can find here low stature grasses, often forming clumps (especially *Nardus stricta* and also *Agrostis capillaris*, *Avenella flexuosa*, *Carex pallescens*, *Festuca ovina*, *Festuca rubra s. lat.*, *Luzula campestris* and *L. sudetica* etc.) and also great amount of herbaceous oligotrophic species (*Campanula brabata*, *Gentiana spp.*, *Galium pumilum*, *Hieracium spp.*, *Homogyne alpina*, *Hypericum maculatum*, *Potentilla aurea*, *Pulsatila scherfelii*, *Rhiantus pulcher* or *Thesium alpinum*) (Chytrý et al., 2001) (fig.12).

This group can be divided into three smaller ones, according to zone where it occurs. Subalpine *Nardus* grasslands can be found around the tree line and can be further divided into

primary *Nardus* grasslands, occurring naturally, and secondary *Nardus* grasslands, occurring on oligotrophic soils after deforestation. Montane *Nardus* grasslands are often combination of alpine species and submontane species. Interesting species of montane *Nardus* grasslands are for example *Campanula bohemica* or *Viola lutea subsp. sudetica*. Last group are submontane *Nardus* grasslands, where we can find for example *Campanula rotundifolia*, *Pedicularis sylvatica* or *Viola canina* (Chytrý et al., 2001).



Fig. 12: Oligotrophic grassland, dominant species – *Avenella flexuosa* (source: author).

These biotopes are the most endangered by eutrophication and abandonment of traditional mowing, locally by forestation. To preserve this biotope regular mowing and occasional grazing is needed (Chytrý et al., 2001).

3.2.2.3 Mesophile grasslands

These regularly mowed or grazed meadows occur on soils with mediocre concentration of nutrients. Mesic *Arrhenatherum* meadows and Montane *Trisetum* meadows are included in this group.

The first biotope is a subalpine type of meadows with dominating low stature grasses – *Agrostis capillaris*, *Anthoxanthum odoratum*, *Festuca rubra s. lat.*, *Trisetum flavescens*, *Dactylis glomerata*, *Holcus lanatus*, *Poa pratensis* etc. Herbaceous plants as *Geranium pratense*, *Trifolium pratense*, *Campanula patula* or *Knautia arvensis* can be also seen often in these meadows. Height of these meadows can reach up to 1 metre, Bryophytes are almost absent. The types with dominating *Festuca rubra s. lat.* are more common in higher altitudes and, regarding environmental protection, these are also the most precious biotopes (Chytrý et al., 2001).

The second biotope of this group – Montane *Trisetum* meadows occur in montane zones from around 600 metres above sea (fig. 13). The upper boundary of the occurrence of these meadows is the tree line. These meadows are characterized by grasses like *Agrostis capillaris*, *Anthoxanthum odoratum s. lat.*, *Festuca rubra s. lat.*, *Phleum rhaeticum*, *Poa chaixii* or *Trisetum flavescens*. The dominant herbaceous species are *Bistorta major*, *Geranium sylvaticum*, *Meum athamanticum*, *Phyteuma nigrum* or *Silene dioica*. The species specific for this biotope in The Giant Mountains are *Campanula bohemica* and *Viola lutea subsp. sudetica* (Chytrý et al., 2001).



Fig. 13: A montane *Trisetum* meadow (source: author).

Over-fertilization and abandonment followed by succession of forests, other montane herbaceous biotopes or invasive plants influence negatively these biotopes. To prevent these negative effects, regular mowing and occasional grazing are necessary. Liming and fertilization is recommended in case of the occurrence of *Nardus* grasslands (Chytrý et al., 2001).

3.2.2.4 Waterlogged grasslands

Waterlogged grasslands can be divided into two groups, depending on the prevailing type of vegetation (grasses versus herbaceous vegetation) (fig.14).



Fig. 14: Waterlogged grasslands. A meadow with dominating *Cirsium rivulare* (on the left) and a meadow with *Filipendula ulmaria* subsp. *ulmaria* and *Geranium palustre* (on the right) (Chytrý et al., 2001).

Wet *Cirsium* meadows are meadows with dominating grasses (*Agrostis canina*, *Carex* spp., *Juncus effusus*, *Poa palustris*, *Scirpus sylvaticus*), however also herbaceous vegetation occur here (*Angelica sylvestris*, *Caltha palustris*, *Trollius altissimus* and of course *Cirsium*

spp.). Also other species from surrounding biotopes can be present, for example species of *Nardus*, peat or mesophile meadows. Bryophytes are not present, except at stands where the process of ulmification occurs. The vegetation in the wet *Cirsium* meadows can withstand short-termed flooding. Drainage, abandonment and overgrowing trees endanger this biotope. The regular mowing is the only way how to preserve it (Chytrý et al., 2001).

Wet *Filipendula* grasslands are often stands with monodominant vegetation, usually *Filipendula ulmaria* subsp. *ulmaria*, *Geranium palustre* or *Lysimachia vulgaris*. However also species from wet *Cirsium* meadows described previously can occur. Specific species present at higher altitudes is *Valeriana excelsa* subsp. *procurrens*. This biotope usually develops from fallow abandoned wet *Cirsium* meadows (Chytrý et al., 2001).

Locally waterlogged parts of other biotopes occurring in the study area with the measure of waterlogging equal or higher than 4 and species *Myosotis palustris*, *Ranunculus spp* or *Deschampsia cespitosa* were also considered as waterlogged grasslands.



Fig. 15: A locally waterlogged area with *Myosotis palustris* (source: author).



Fig. 16: Degraded grasslands – dominating species: *Calamagrostis villosa* (upper left), *Holcus mollis* (upper middle), *Carex brizoides* (upper right), *Deschampsia cespitosa* (lower left), *Rumex alpinus* (lower middle) and *Veratrum album* (lower right) (source: author).

3.2.2.5 Degraded grasslands dominated by *Poales* (except *Nardus stricta*)

The grassy stands (mostly oligotrophic) damaged by wrong or none management belong to this category. Under none management, these meadows undergo natural succession. This means that some of species become more abundant than others, diversity decreases and meadows become mono dominant. Other characteristic features of these degraded grasslands are the cumulation of old biomass and the absence of traditional meadow species (Krahulec et al., 1997).

The species usually dominating these grasslands are: *Calamagrostis villosa*, *Holcus mollis*, *Deschampsia cespitosa* and *Carex spp.* in more waterlogged areas (fig. 16).

3.2.2.6 Degraded grasslands dominated by *Dicotyledons*

As in the previous class, these grassy stands have been damaged by wrong or insufficient management. Again these grasslands show decreased species diversity in comparison with their non-degraded counterparts and they also lack traditional meadow species. However as these biotopes usually occur on soils richer in nutrients (naturally or as a result of over-fertilization), different species become dominant here (Krahulec et al., 1997).

The most prevailing species are *Rumex alpinus* and *Veratrum album* (fig.16).

3.2.2.7 Subalpine *Vaccinium* vegetation

Subalpine *Vaccinium* vegetation occurs on windward and leeward sides of the mountain slopes at altitudes close to the tree line. The biotope is represented especially by low scrubs of *Vaccinium myrtillus*, less often also by *Vaccinium vitis-idaea* species. The main species occur often in mosaic with juvenile woody plants (e.g. *Picea abies*, *Pinus mugo* or *Sorbus aucuparia subsp. glabrata*) or grasses (*Calamagrostis villosa* or *Avenella flexuosa*). Herbaceous species like *Gentiana asclepiadea*, *Homogyne alpina*, *Melampyrum pratense* or *Trientalis europaea* are also common. Eutrophication and trampling by tourists and skiers have the most negative effect on *Vaccinium* scrubs (Chytrý et al., 2001).

Secondary submontane and montane heathlands also belong to this category. This category encompasses secondary heathlands which developed as a result of deforestation of acidophilous beech forests or montane spruce forests. *Calluna vulgaris* together with already mentioned *Vaccinium* species occur here. Other present species are e.g. *Solidago virgaurea subsp. virgaurea*, *Potentilla erecta*, Bryophytes and lichens. Common management performed on these biotopes is removing of self-seeding woody plants and grazing (Chytrý et al., 2001).

3.2.2.8 Springs

Meadow springs without tufa formation and also subalpine springs belong to this class.

Meadow springs are springs with short-stemmed herbaceous or mossy vegetation with dominating *Montia* spp. or *Stellaria uliginosa*. Other common species of vegetation are e.g. *Agrostis stolonifera*, *Glyceria fluitans*, *Holcus* spp., *Poa trivialis*, *Carex canescens* and *Carex nigra*. *Cardamine amara* or *Equisetum fluviatile* can be also found here. Bryophytes (*Rhizomnium punctatum*, *Philonotis Fontana*, *Calligeron* spp. or *Bryum* spp.) can create continuous vegetation cover or they are suppressed by *Montia* spp. and *Stellaria* spp. Meadow springs are endangered by eutrophication, drainage or mechanical destruction and it is important to maintain non-forested areas in their surroundings for their preservation (Chytrý et al., 2001).

Subalpine springs are non-continuous type of springs usually occurring above tree line. Both, herbaceous vegetation and Bryophytes can be dominating. From herbaceous species *Carex nigra*, *Trichophorum alpinum*, *Allium schoenoprasum* subsp. *alpinum*, *Epilobium alsinifolium* or *Swertia perennis* can be found in this biotope. Common Bryophytes are *Cratoneuron commutatum*, *Dicranella palustris* and *Brachythecium*, *Bryum* and *Philonotis* spp. Subalpine springs are endangered by the same factors as the previously described spring biotope (Chytrý et al., 2001).

3.2.2.9 Subalpine tall-fern vegetation

This biotope is present at wind-protected spots above the tree line, on the slopes and cirques and shadowed spots at or under the tree line. The dominant plants are *Athyrium distentifolium*, *Dryopteris filix-mas* and less often *Athyrium filix-femina*. The dominant species are complemented by herbaceous vegetation like *Aconitum callibotryon*, *Adenostyles alliariae*, *Cicerbita alpine* or *Veratrum album* subsp. *lobelianum*. The vegetation is usually 80 to 120 cm tall and closed. Bryophytes are not common because of abundant old fern vegetation (Chytrý et al., 2001).

4. Data

4.1 Aerial imagery

The data utilized in this master thesis, is hyperspectral aerial data acquired by AISA DUAL. Scanning and pre-processing of data was performed by Geodis (Czech Republic) (AISA, 2013).

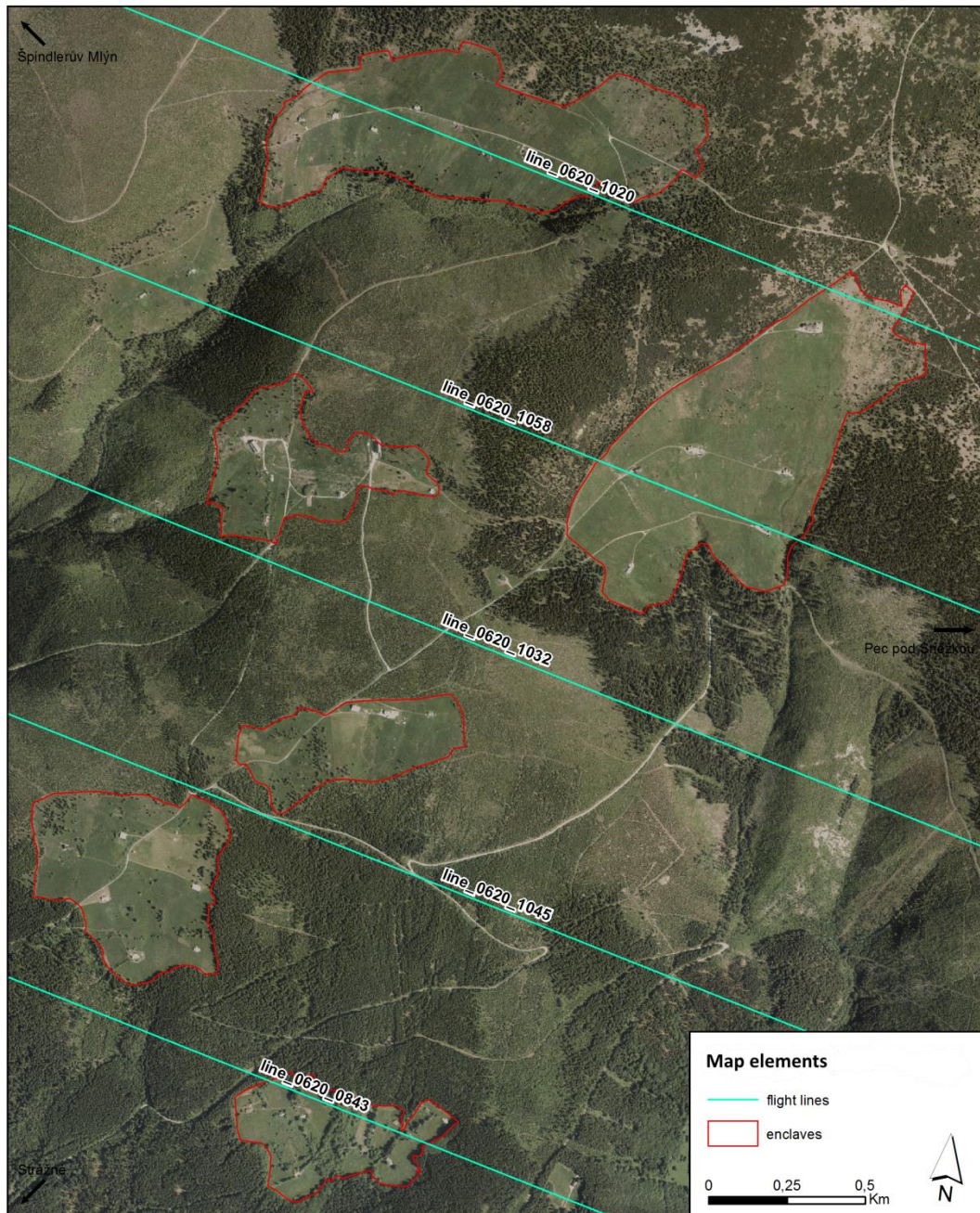


Fig.17: A sketch of utilized flight lines (source: own ArcGIS output). First four digits mean the date of acquisition (in format MMDD); last four digits mean the time of acquisition (in format HHMM, before noon).

Five flight lines were used for the classification of the vegetation in the study area (fig.17). All these lines were recorded on 20th June 2013. The time of acquisition was between 8:45 am and 11:00 am, which is not appropriate time of recording as the sun is still very low on the horizon casting long shadows. This problem will be discussed later in the chapter 6.1.3. Cloud cover during acquisition was reported as 13% (AISA 2013).

The images were taken in 498 spectral bands (244 spectral bands for AISA EAGLE and 254 spectral bands for AISA HAWK) covering wavelengths from 400nm to 2500 nm (AISA DUAL, datasheet). Pixels in AISA HAWK images were resampled from 3m spatial resolution to 1m spatial resolution in order to match the spatial resolution of AISA EAGLE images.

4.1.1 The description of utilized AISA sensors

4.1.1.1 AISA DUAL

AISA DUAL is a high-performance airborne hyperspectral sensor for simultaneous acquisition of images in VNIR and SWIR part of spectrum (fig. 18). It is comprised of two airborne pushbroom hyperspectral sensors AISA EAGLE and AISA HAWK and therefore able to record data in 498 bands at wavelengths from approximately 400 nm to 2500 nm. AISA DUAL sensor was employed in dual computer data acquisition setup (AISA DUAL, datasheet).

The comparison of AISA EAGLE and AISA HAWK sensors is in table 3.

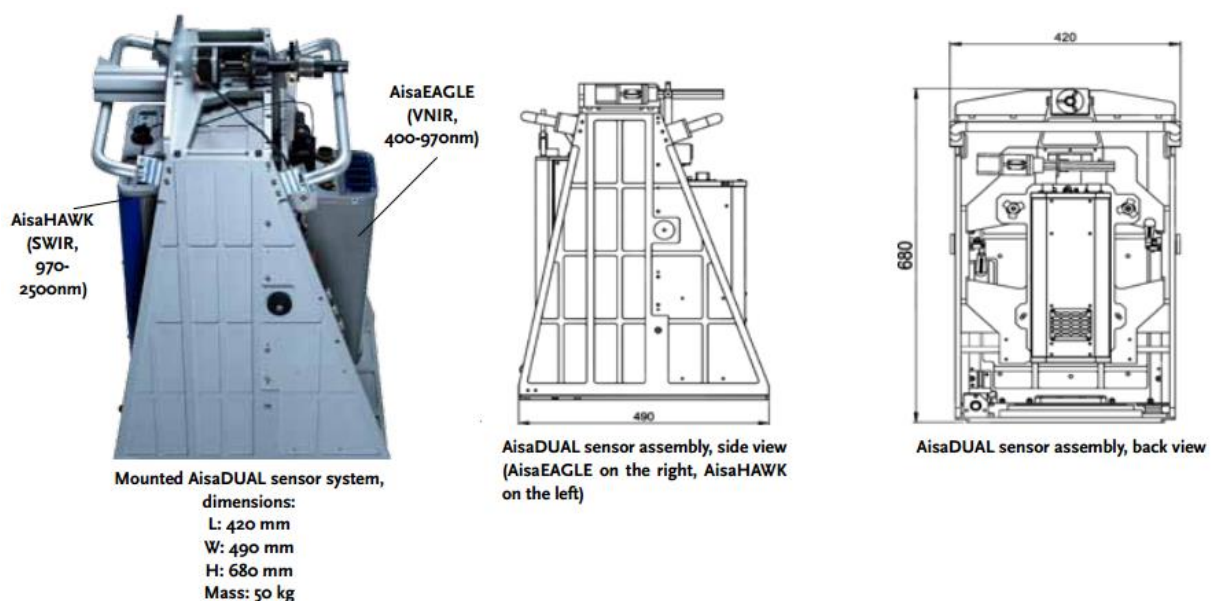


Fig.18: AISA DUAL sensor (AISA DUAL, datasheet).

4.1.1.2 AISA EAGLE

AISA EAGLE is airborne VNIR pushbroom hyperspectral sensor recording up to 488 spectral bands at wavelengths from 400 nm to 970 nm (fig.19). Sensor is part of a bigger system including GPS/IMU, a unit recording data and a robust computer to store data. AISA EAGLE sensors have been part of many applications, such as forestry management, vegetation cultivation, environmental investigations, precision farming, target identification, water assessment and land use planning (AISA EAGLE, datasheet).

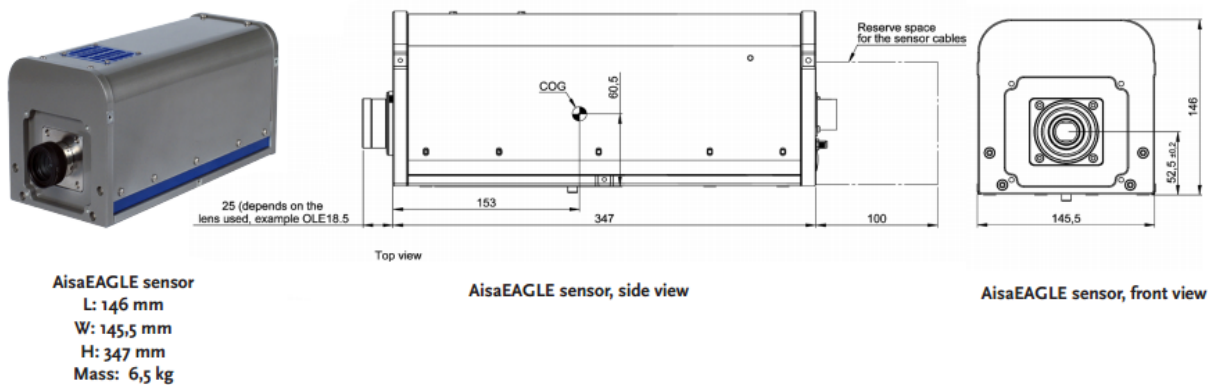


Fig.19: AISA EAGLE sensor (AISA EAGLE, datasheet).

Table 3: A comparison of AISA EAGLE and AISA HAWK sensors. The information about sensors was gained from their datasheets. Abbreviation CCD means charge-coupled devices this type of detector is often used in pushbroom sensors (Jones and Vaughan, 2010). MCT means mercury-cadmium-tellurium and is the only common material for detecting IR radiation (Norton, 2002).

Sensor	AISA EAGLE	AISA HAWK
Type of sensor	pushbroom	pushbroom
Detector	CCD matrix	MCT matrix with cooler
FOV (degrees)	29.9 / 37.7	17.8 / 24.0 / 35.5
Ground resolution at 1000 m alt. (FOV dependent; m)	0.52 / 0.68	0.97 / 1.34 / 2.0
Spectral range (nm)	400 - 970	970 - 2450
Spectral resolution (nm)	3.3	8.5
Number of spectral bands	488 / 244 / 122 / 60	254
Spectral sampling (bands dependent; nm)	1.15 / 2.3 / 4.6	5.8
Output	12 bits digital	14 bits digital
SNR	1250:1 (maximum theoretical)	800:1 (peak)
Storage temperature (°C)	- 20 ... + 50	- 20 ... + 50
Operating temperature (°C)	+ 5 ... + 40 (non-condensing)	+ 5 ... + 40 (non-condensing)

4.1.1.3 AISA HAWK

AISA HAWK is the first maintenance free airborne SWIR hyperspectral sensor on the market (fig.20). Its biggest advantage is the distinctive size as the AISA HAWK is also the smallest of all the airborne SWIR hyperspectral sensors. AISA HAWK sensors are utilized in applications where characteristic signatures of chemical compounds are necessary and VNIR range is not sufficient (AISA HAWK, datasheet).

Sensors for acquiring images at SWIR wavelengths are highly influenced by temperature. Therefore AISA HAWK sensors have also in-built coolers to prevent changes in temperature and consequent lower quality of data even on long flights (AISA HAWK, datasheet).

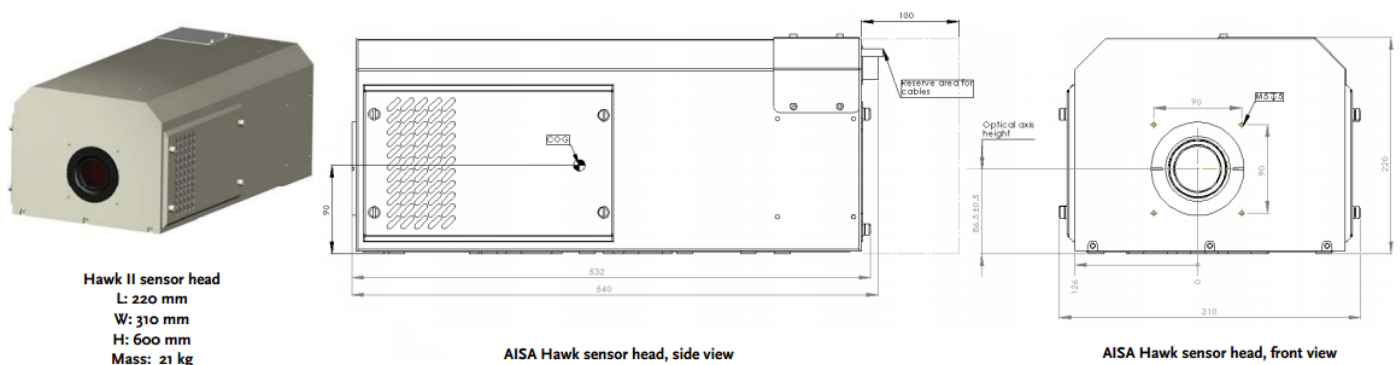


Fig.20: AISA HAWK sensor (AISA HAWK, datasheet).

4.1.2 Pre-processing of aerial imagery

The reflected signal from surface is usually recorded as a simple digital number. Digital number describes intensity of the radiation in a particular wavelength range at a sensor and can be converted to radiance at the sensor (fig.21). However for majority of remote sensing applications reflectance at surface is needed. Therefore imagery needs to be corrected for atmospheric absorption and the scattering of electromagnetic radiation travelling between the sun, the surface and the sensor (Jones and Vaughan, 2010).

Furthermore data are corrected for radiometric and geometric errors, e.g. for errors caused by faulty detectors (banding or stripping), non-linear scanning and response of detectors, motion of the platform on which sensors are mounted and panoramic distortion (a “bowtie” effect). Geometric corrections are of high importance especially when airborne data is utilized, because such data is sensitive to changes in attitude (yaw, roll and pitch) and altitude (Jones and Vaughan, 2010). Regarding hyperspectral pushbroom sensors one has to be also aware of

spectral distortion, called “spectral smile”, caused by a shift in wavelength in the across-track direction of the sensor (Imaging spectroscopy, 2013).

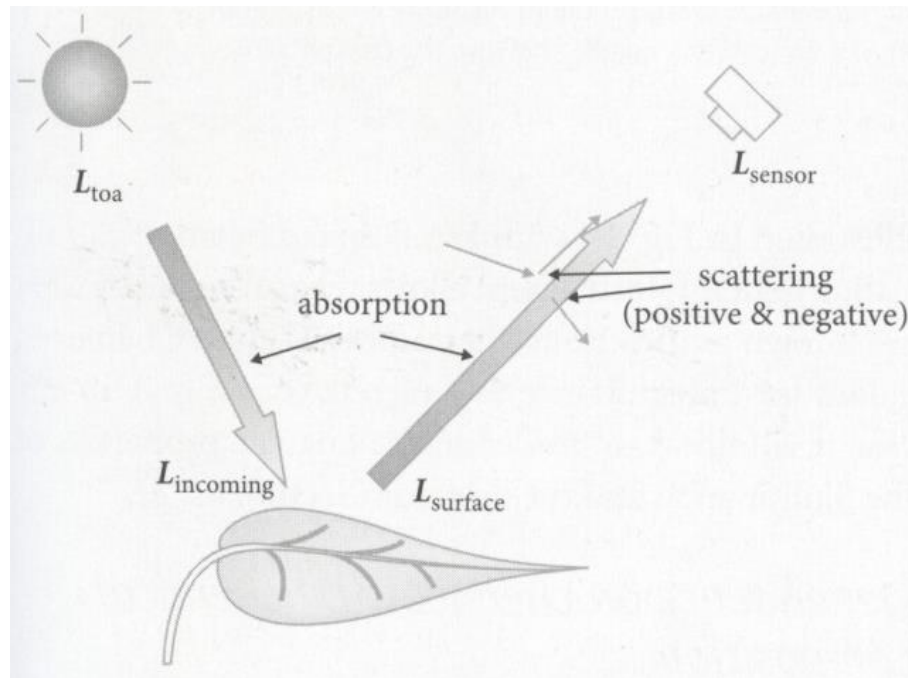


Fig. 21: An illustration of the relation between the radiance received at a sensor (L_{SENSOR}), the radiance reflected from a surface ($L_{SURFACE}$) and the incoming radiances at the top of the atmosphere (L_{TOA}) and at the surface ($L_{INCOMING}$) (Jones and Vaughan, 2010).

If a true geometric representation is required, the image must be warped (geo-located) so that it will correspond exactly with map coordinates or with other images, this is achieved by using a series of ground control points (GCP) in the image and on the map (Jones and Vaughan, 2010).

4.1.2.1 Pre-processing in GEODIS

Specialized software CaliGeoPro and ENVI/IDL was utilized for pre-processing of AISA aerial imagery. Spectral properties of calibration areas were measured by ASD field spectrometers (AISA, 2013).

Hyperspectral imagery was calibrated radiometrically using measured spectral characteristics of small areas with uniform reflectance properties. Imagery was also corrected for geometric distortions (fig.22). Atmospheric corrections were performed on radiometric and geometric corrected data, also using measured characteristics from the field (AISA,

2013). The result of pre-processing in GEODIS was level 2 product (geo-coded, radiometrically, geometrically and atmospherically corrected imagery).

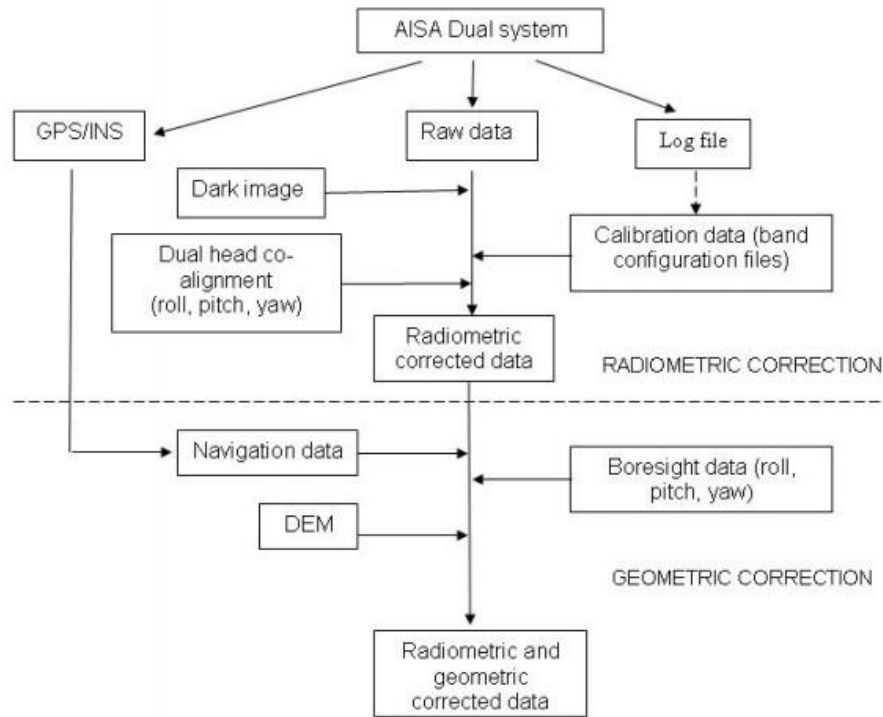


Fig.22: Workflow for pre-processing of hyperspectral imagery in GEODIS (AISA, 2013).

4.1.2.2 Layer stacking and mosaicking

Upon receiving the data from GEODIS, its volume was lowered by multiplying the reflectance values by 10000 and converting the data format from double to unsigned integer, because this type of a number needs less space to store (16-bit for unsigned integer versus 64-bit for double). As the dual computer data acquisition setup was utilized during hyperspectral imaging, it was necessary to put together spectra from both sensors AISA EAGLE and AISA HAWK. This was done by *Layer Stacking* tool in SW ENVI 4.7 with following parameters: x pixel size and y pixel size was set as 1 m and resampling method was chosen as nearest neighbour. The layers were stacked only in the area of interest (under spatial subset) and also spectral subset was defined. While spatial subset was needed because of the huge file sizes (about 200 GB per line per sensor) and computational burden, spectral sub-setting was performed because of spectra overlap (fig.23). The bands to be preserved were chosen in a way so the majority of bands from AISA EAGLE sensor were kept, because AISA EAGLE

has better spatial and spectral resolution. Therefore bands number 1 to 252 of AISA EAGLE and bands number 4 to 244 of AISA HAWK were kept for further processing.

Based on previous bad experience with mosaicked images (consultations with the supervisor) and on significant differences in brightness values between lines, it was decided that lines would be processed separately without mosaicking.

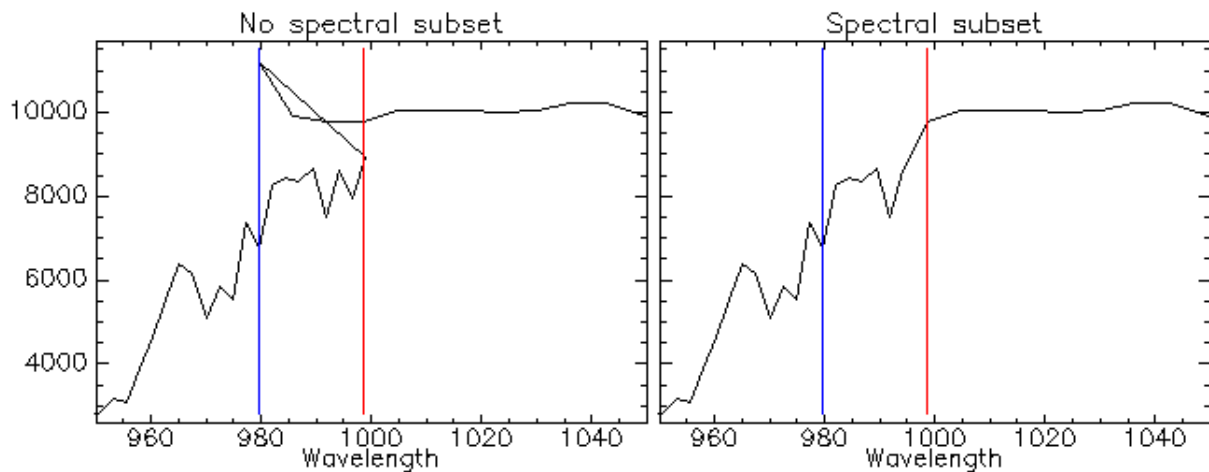


Fig. 23: Spectra overlap of AISA EAGLE and AISA HAWK sensors (source: own ENVI analysis). The area between blue and red lines is the area of overlap. The image on the left shows spectrum of a particular pixel after layer stacking without spectral subset, the image on the right shows spectrum of the same pixel after layer stacking with spectral subset.

4.1.2.3 Principal component analysis

The PCA was performed in order to decrease computational demand on hardware and also to get rid of noisy bands. The traditional PCA method was utilized as described in chapter 3.1.1. SW ENVI 4.7 and *Principal Components / Forward PC Rotation* tool was used for the analysis. Based on the result of the PCA analysis, first seven bands of the new dataset actually carried any information and thus these bands were used in further pre-processing and for the experiments.

4.1.2.4 Masking of non-classified classes

To fully use the potential of hyperspectral data in the classification of meadows, masks for shadowed areas, for areas outside of the area of interest and for legend classes A1 to A5 were created. Masks were built in several-step process including:

- classifying the PCA images using unsupervised IsoData classification to automatically distinguish land cover types to be masked
- making additional polygons in SW ArcGIS 10.2 for areas which were not identified by IsoData classifier (e.g. some roads and deciduous trees)
- building masks by *Build Mask* tool in SW ENVI 4.7
- applying masks on the PCA processed imagery by *Apply Mask* tool in SW ENVI 4.7

IsoData classifier was run with the following specifications: number of classes was set from 10 to 50; maximum number of iterations was set to 100 and every other parameter was left as default. The unsupervised classification produced usually around 30 – 50 classes, these were combined into meaningful groups using *Class colour mapping* and *Combine classes* tools in SW ENVI 4.7. As it was mentioned before, it was necessary to create additional polygons for some roads and deciduous trees. This was done in ArcGIS 10.2 utilizing tools in *Editor*. Polygons were mapped according to orthophoto maps, which are available via WMS services of State Administration of Land Surveying and Cadastre. Orthophotos were used because of their high spatial resolution and timeliness (ČÚZK, orthophoto).

The third step was to build actual masks. Masks were built in SW ENVI 4.7 in two-step procedure. First of them, with IsoData classified rasters (range set to classes to be masked) and shapefiles of additional polygons of roads and trees as attributes and using selected areas “off” and logical “OR”, resulted in a sub-mask showing only meadows. The sub-mask was further processed in the second step, where it was subset to contain only the area of interest.

Finally masks were applied to the PCA images making them ready for the experiments.

4.1.3 Errors in the aerial data

Some errors in the aerial data were found during its pre-processing.

The major drawback of several lines of the hyperspectral imagery is the time of acquisition. The line where Husí Boudy and part of Lahrový Boudy are situated was imaged at around 8:45 am. Therefore huge shadows are present in these images, which make many

pixels impossible to use for the analysis (fig. 24). It would be convenient in the future to avoid acquiring images during such early hours of the day.

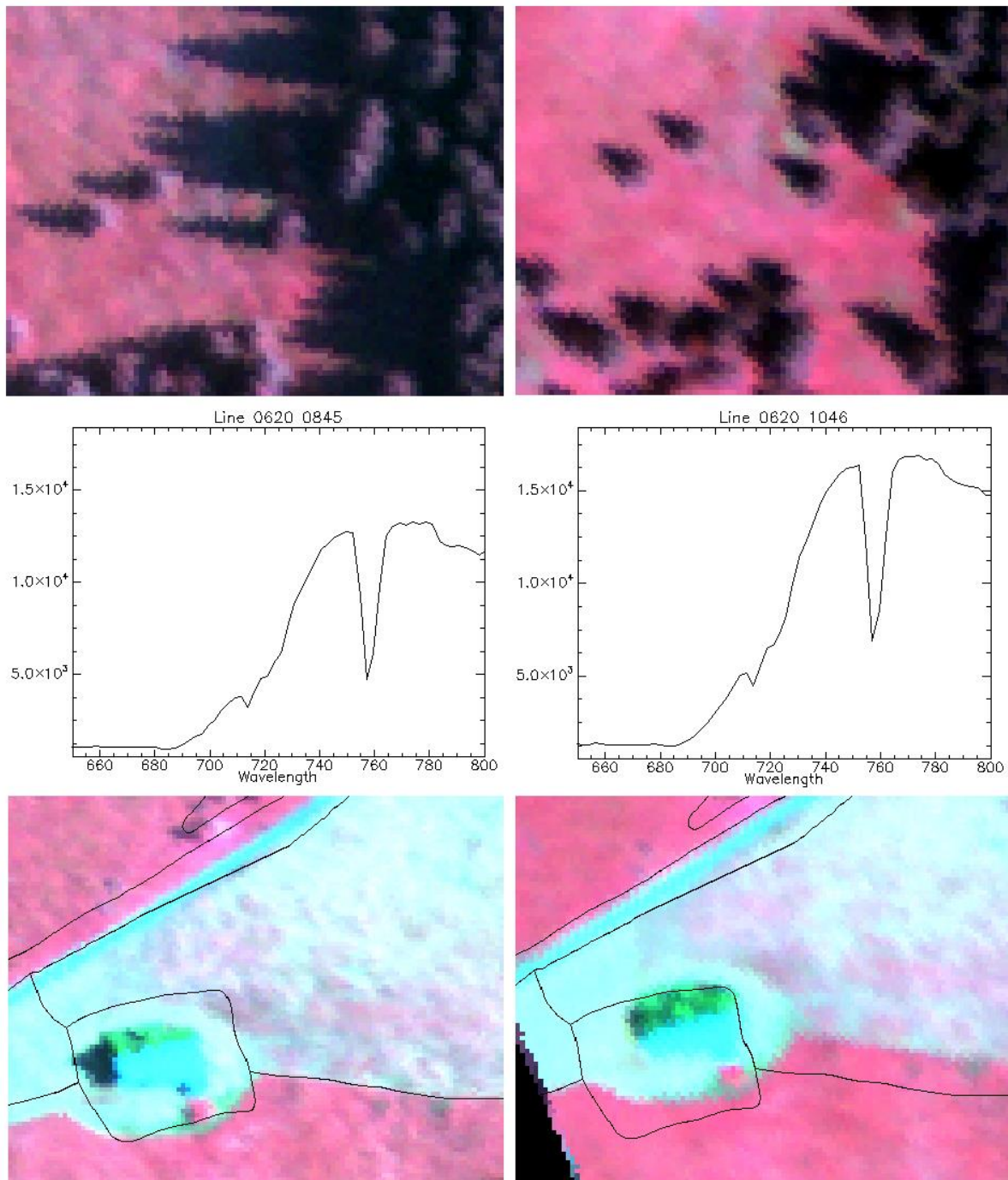


Fig. 24: Errors in the aerial imagery, example from the location Lahrový Boudy (source: own ENVI analysis). The upper images show shadows due to early daytime of acquisition (line_0620_0845 on the left, line_0620_1046 on the right). Obviously there are more shadows in the imagery which was recorded earlier (8:45am versus 10:46am). The plots show differences in brightness values of the same pixel. The lower images show geometric distortions on the edge of a line. A vector layer was added to better illustrate the distortions. The imagery on the left (line_0620_0845) is considered as not distorted.

Furthermore, there have been observed differences in brightness between lines (fig. 24). In fig. 24 can be clearly seen that same places have different brightness depending on the line. This suggests that the radiometric or the atmospheric corrections were not performed properly or that wrong areas were chosen for calibration measurements. However as the experiments do not include deriving the qualitative information from the imagery, this error should not have strong impact on the results of experiment.

Moreover, geometric distortions and misplacing of objects can be observed as one moves towards the edge of the line (fig. 24, lower images). These errors may result in inaccurate classifications as the spectral information of a misplaced pixel may not be of the same vegetation as it was mapped in the field. Nevertheless, such distortions are common on the edges of an image and are due to lower amount of GCPs available (Campbell and Wynne, 2011). One solution for this problem may be to intentionally collect more GCP at the spots where the known edges of lines are.

At last, it was learnt, that some enclaves (Lahrový Boudy and Zadní Rennerovky) were not whole on any of available flight lines. This is not error in the true sense of the word, however as the lines were processed separately, this fact hinders the quality of analyses on these enclaves.

Given all the mentioned errors and shortcomings in the aerial imagery, the enclaves Přední Rennerovky and Friesový Boudy, where these were of the lowest magnitude, were chosen as acceptable for the experiments.

4.2 Field mapping

Field mapping was realized in July and August 2014. The mapped polygons were drawn in the printed orthophoto maps and simultaneously parameters of interest of the vegetation in the polygons were noted into tables. The recorded parameters were, for example, a measure of waterlogging of soil, a type of vegetation and height of vegetation. In the end, mapped vegetation was categorized into classes according to the legend in chapter 5.2.1.

4.2.1 Pre-processing of the field data

The field data were pre-processed using SW Microsoft Access, Microsoft Excel, ArcGIS 10.2 and Python scripts (module ArcPy).

The drawn polygons had to be transformed into shapefiles. This was done using *ArcScan* tool. In order to be able to use this tool, lines of polygons were first redrawn on transparent foils and scanned into digital form (raster format). Shapefiles were then also georeferenced utilizing standard ArcGIS methods.

As the next step, description of vegetation classes had to be linked with the polygons in shapefiles. A Python script was written to perform this task (appendix I). During this step georeferenced vector layers, containing information on classes, were also divided into several smaller vector layers in the way that the polygons of each class occurring on the location were in one vector layer. Furthermore a buffer with the distance value of -2 was applied to polygons to deal with possible mapping inaccuracies (chapter 6.2.2) and geometrical distortions of the aerial images (chapter 6.1.3).

A fishnet was then created (snapped to the aerial image of a particular location) for each location and also the masks were converted to vector layers and only the areas to be classified were selected in these vector layers. After that, selection by location was made in the fishnet so only the cells of the fishnet being within buffered polygons were selected. This selection was then intersected with the selection made in the mask vector layer and saved as a new shapefile.

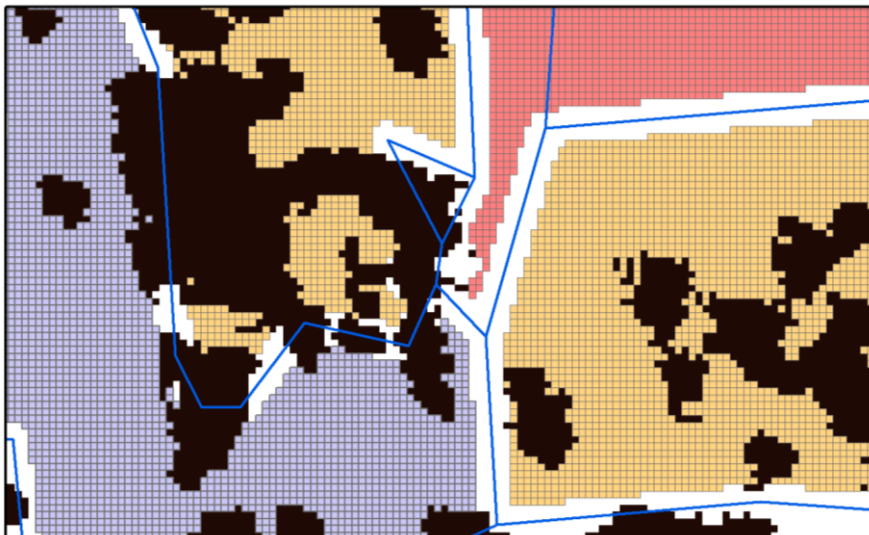


Fig. 25: An example of a set of pixels prepared for the experiments. The red, ochre and blue pixels are the pixels of classes 1 to 3 ready for the selection of training and validating pixels, the blue lines depict field mapping polygons and the black pixels are the masked pixels (source: own ArcGIS analysis).

The intersection step might seem as unnecessary because masks were already applied to the aerial images, but as the experiments also involve tests with amount of training pixels, it is needed to define precisely the area from which the training pixels may be selected. Since the situation could happen, where pixels would be chosen in the masked area and later ignored by the classifier, or worse the spectral signature of the mask would be taken as training one for a particular vegetation class. At this point the field mapping data was prepared for the next stage – defining of training and validating pixels (fig. 25). The form of squared fishnet cells was chosen because it resembles the most the shape of the pixels in the aerial images.

4.2.2 Problems with the field data

As a botanist sees the vegetation communities in a different way than a remote sensing scientist, it is always difficult to reach a compromise. Challenges might occur when creating a legend in a meaningful way for both, the botanist and the remote sensing scientist during a field mapping. Furthermore the field mapping for this master thesis has been done by a botanist hired by KRNAP and a remote sensing scientist was present only during the mapping of few locations. Moreover, the mapping was performed also as a source of information for another, entomological analysis. Therefore the result of the field mapping is a kind of “hybrid” between what is suitable for remote sensing and what is suitable for entomologists and thus it presents several disadvantages for a remote sensing analysis.

At first, as the field mapping was performed without GPS, only by simple drawing polygons into orthophotos, there were some spatial discrepancies between the reality and drawn polygons (fig. 26). Moreover, these discrepancies could have been caused also during georeferencing by not having accurate points of known coordinates in the image. The positions of the objects that were separable using the orthophotos provided by ČÚZK were fixed in SW ArcGIS 10.2 during the pre-processing of the field data. However the different types of meadows could not be distinguished in the orthophotos and therefore it was impossible to fix their positions. Consequently it is assumed that the accuracy of the classification is lower than what the potential of hyperspectral data is.

Secondly, the field mapping was not performed in such a detail as it would be necessary for the hyperspectral data with 1 m spatial resolution. The example is in the fig. 27. The big polygon has been described as a huge pasture with dominating *Deschampsia*

cespitosa, locally waterlogged and occupied by *Rumex alpinus* and it has been categorized as class 6 – degraded meadows with dominating grasses. Such a description is appropriate for entomologists, but not for a remote sensing analysis, as it can be seen also other classes has been detected by a classifier and assigned to a different class, than what the polygon had been categorized to. It is clear that it is not possible to map all the pixels correctly as they might not be distinguishable in the field. Nevertheless, at least the most significant parts with *Rumex alpinus* and the waterlogged parts could have been mapped individually as they were also described separately and they cover substantial part of the polygon. We can also observe in the fig. 30 that as a result of the coarse field mapping, also the lines of the outlying polygon are coarser in comparison with the imaginary lines of the polygons detected by a classifier.



Fig. 26: Mapping inaccuracies (example from the location of Friesovy boudy). Red lines show the inaccurate lines from the field mapping, yellow lines show the fixed lines (missing polygons or shifted polygons) and the orange lines are the correct lines from the field mapping (or those which were used also in the fixed polygons) (source: own ArcGIS output).

In order to properly check the influence of the assessed factors (the amount and the sampling design of training pixels, the choice of a kernel and the setting of other SVM parameters) on the accuracy of classification, a new set of spectrally uniform polygons,

assigned to classes with a certainty, was needed. Therefore the unsupervised IsoData classification was performed on the masked images of Přední Rennerovky and Friesovy Boudy and these results were checked in the field. The spectrally uniform polygons which were classified accurately were marked and the new set of polygons was created according to these. Based on the new field mapping, the location Přední Rennerovky was chosen as the model one and the experiments were conducted on this location.

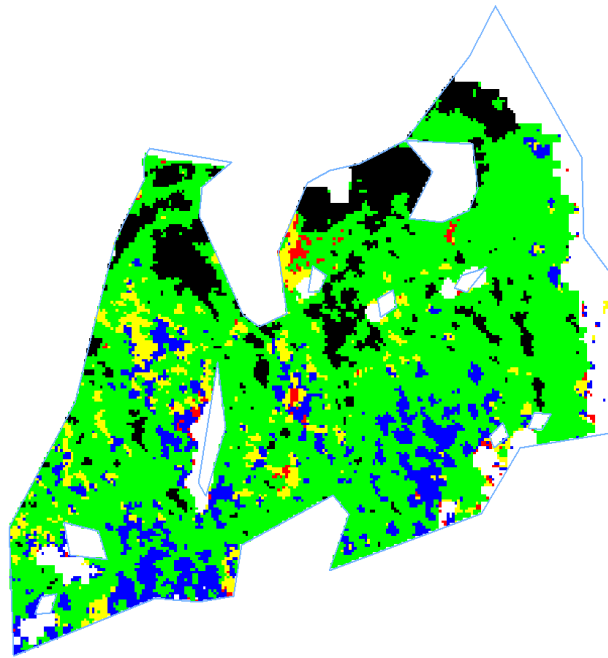


Fig.27: An example of not enough detailed field mapping. All (or maximum) pixels in the polygon should have green colour (class 6 – degraded meadows with dominating grasses). However also pixels of *R.alpinus* (class 5, black), waterlogged grasslands (class 4, blue), oligotrophic grasslands (class 2, yellow) and mesophile grasslands (class 3, red) were detected (source: own ArcGIS and ENVI analysis).

5. Methods

From the previous descriptions of methods currently used for classification of grassland vegetation (chapter 2.3) it can be seen, that the gradient modelling is probably the best method in terms of proximity to real world relationships (Magiera, 2013; Schmidlein and Sassin, 2004). However this method is extremely demanding, for both measuring vegetation or ancillary data (environmental variables) and its computational burden. Therefore it would be too time-consuming and maybe even impossible to classify the whole study area using this method. On the other hand, methods as NN or SVM are quite easily applicable, with appropriate properties for classifying hyperspectral data, while they have shown satisfactory accuracies (table 4) and thus seem as the most suitable for this task.

Table 4: The examples of the studies comparing performance of SVM and NN classifier (source: author).

Biotope classified	Data used	SVM accuracy (highest achieved)	NN accuracy (highest achieved)	Reference
Grasslands and woodlands in the eastern Maryland (USA)	Landsat TM	75%	74%	Huang et al., 2002
Crops in Barrax (Spain)	HyMap	96%	94%	Camps-Valls et al., 2004
Indiana Pines (northwest Indiana)	AVIRIS	94%	88%	Melgani and Bruzzone, 2004
Crops around Littleport (eastern England) and crops and pastures near La Mancha Alta (central Spain)	Landsat ETM+ and DAIS	87.5% and 94%	85% and 90%	Pal and Mather, 2005
Rural area in the South-West Florida	Landsat TM	79.2%	78.4%	Dixon and Candade, 2008
Tropical forest of Shoolpaneshwar Wildlife Sanctuary	Hyperion	70.7%	80.5%	Vyas et al., 2011
Mediterranean vegetation of Attica and Voiotia prefectures (central Greece)	Hyperion	89.26%	85.95%	Petropoulos et al., 2012

As it can be seen accuracies achieved by both classifiers in different studies vary from under 80% to over 90%, mostly depending on data used (multispectral versus hyperspectral), spatial resolution and complexity of classified classes. However in 6 out of 7 studies SVM classifier outperformed NN, regarding accuracy and computational time. Authors attributed SVM's higher performance to several reasons, but the major reason stated was, that SVM is able to identify an optimal separating hyperplane to distinguish between different vegetation communities, even when limited amount of training samples is available. On the contrary NN classifier may not be able to place such a decision boundary under the given circumstances

(Petropoulos et al., 2012; Melgani and Bruzzone, 2004, Huang et al., 2002). In addition, training a neural network requires the tuning of several parameters (activation function, learning rate, network architecture, training algorithm, stopping criterion and others), while for the training of SVM only few parameters have to be selected (the type of kernel, γ parameter and C parameter) (Camps-Valls et al., 2004) and in the case of polynomial kernels also the degree of polynomial (Huang et al., 2002). The problem of kernel choice is at the moment the biggest drawback of the SVM classifier, because there are not any general guides in the literature about the suitable selection of kernels for particular applications (Mountrakis et al., 2011).

Based on the studies mentioned in the table 4 (where in majority SVM outperformed NN) and also based on the properties of the SVM classifier (only a few parameters to tune and faster computational times), it has been chosen as the main classifier to experiment with. Nevertheless, NN classifier has shown satisfactory results too, therefore it has been chosen as the reference classifier.

The experiments with the SVM classifier included tuning SVM parameters (the type of kernel, γ parameter, C parameter and in the case of the polynomial kernel – the degree of polynomial) and experimenting with the size of training dataset and the spatial sampling of training pixels. At first, the SVM parameters were defined using certain amount and spatial sampling design of training dataset and then the rest of the amount–sampling design combinations of the training dataset were investigated. The random sampling design was used as the initial one for tuning parameters, because according to Huang et al., the pixels collected using this sampling design are not spatially correlated and the spectral variability of each class is not underestimated (Huang et al., 2002). The initial amount of pixels to tune parameters with was chosen based on the studied literature and HW requirements as 200 for each class (more information in chapter 5.2.1).

All the experiments and classifications were performed in SW ENVI 4.7.

5.1 Tuning parameters of the SVM classifier (step 1)

First of all, the different types of kernel functions were tested (linear, RBF and all types of polynomial kernels offered by ENVI). Furthermore several values of γ parameter for RBF kernel and the best performing polynomial kernel were tested. In the end the range of penalty parameter values was tested for the best combination of RBF kernel and its γ

parameter and also for the best performing polynomial kernel and its γ parameter. The classification probability threshold value was set to default (0.0) in all the cases, forcing all pixels to be classified.

5.1.1 The penalty parameter

The penalty parameter (or C parameter) defines the magnitude of allowed errors (Camps-Valls et al., 2004), in other words it lets you control the trade-off between allowing training errors and forcing rigid margins (ENVI, SVM background). A high value of the penalty parameter forces the SVM classifier to avoid classification errors (Dixon and Candade, 2008) and creates more accurate model, such model however might not generalize well (ENVI, SVM background).

In ENVI this value is a floating-point value greater than 0.01, the default value is 100.0 (ENVI, SVM background). The penalty parameter was tested between 0.01 and 100 with the increment of 10. The values of the penalty parameter to be tested were determined based on the studied literature (table 5) with the emphasis on the values which were used by the scientists working with ENVI (e.g. Marcinkowska et al., 2014; Petropoulos et al., 2012; Vyas et al., 2011).

Table 5: The summary of the kernels, the values of the penalty parameter and the amount of training pixels utilized in the studied literature. The number in the brackets after the polynomial kernel notes its degree. The symbol “-” means that the information was not mentioned in the cited article (source: author).

Type of the kernel used	Values of the penalty parameter used	Number of training pixels used for each class	Reference
RBF	2 - 100	42, 84 and 147 pixels	Belousov et al., 2002
RBF, polynomial (1 - 4)	-	2 – 20 pixels and 2 – 20 % of pixels	Huang et al., 2002
RBF, polynomial (1 - 8)	1 - 100	150 pixels	Camps-Valls et al., 2004
RBF	16	75 pixels	Foody and Mathur, 2004
linear, RBF	1 - 100	236 – 1245 pixels	Melgani and Bruzzone, 2004
RBF	5000	100 pixels	Pal and Mather, 2005
RBF, polynomial (3)	1000	1 % of pixels	Dixon and Candade, 2008
RBF	10 - 1000	95 – 1434 pixels	Demir et al., 2009
linear, RBF, polynomial (-)	-	-	Vyas et al., 2011
RBF	1.4 – 8.3	16 – 188 pixels	Chan et al., 2012
RBF	100	-	Petropoulos et al., 2012
linear	-	400 pixels	Marcinkowska et al., 2014

5.1.2 The choice of a kernel

Chan et al and Vyas et al. have tested the accuracy of classification using different kernels. According to Chan et al. the linear kernel gave poor accuracies (Chan et al, 2012), which is in agreement with results of Vyas et al (Vyas et al, 2011). Furthermore Chan et al stated that although polynomial kernels gave satisfactory accuracies they were too time-consuming with three parameters to tune, so the kernel of their choice was Gaussian because it produced quite high accuracies while computing time was shorter than for polynomial kernels (Chan et al., 2012). Vyas et al used only accuracy criteria for choosing kernel, but also in their case the Gaussian kernel gave the best accuracy results (Vyas et al., 2011). Petropoulos et al chose as well Gaussian kernel, but their decision was based on the fact that it requires the definition of only one parameter to run (Petropoulos et al., 2012). On the contrary Huang et al report that polynomial kernels gave slightly higher accuracies than Gaussian kernels, however they also admitted that as order of polynomial kernel increases the computational time for classifier increases even more rapidly (Huang et al., 2002). Marcinkowska et al. report that the linear kernel gave the best results, which is quite the opposite from other studies (Marcinkowska et al., 2014).

SW ENVI offers four types of kernels: linear, RBF (Gaussian), sigmoid and polynomial. From those, only three (linear, RBF and polynomial) were tested in the experiments because those gave the reasonable accuracies according to the references (table 5). The mathematical representation of the utilized kernels is as follows (ENVI, SVM background):

$$\begin{aligned} \text{Linear kernel:} & \quad K(x_i, x_j) = x_i^T x_j; \\ \text{RBF kernel:} & \quad K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0; \\ \text{Polynomial kernel:} & \quad K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0; \end{aligned}$$

where:

γ is the gamma parameter in the kernel function

d is the polynomial degree in the kernel function

r is the bias term in the kernel function.

Furthermore all degrees possible to choose for the polynomial kernel (1 to 6) were tested. The gamma parameter was tested in the whole range selectable in ENVI 4.7 – from 0.01 to 10000 with increment of one order on a logarithmic scale. The logarithmic scale and the increment were chosen especially because of the wide range of the selectable values. The ENVI default value of the gamma parameter is the inverse of the number of bands in the input image (ENVI, SVM).

5.2 Defining a training and a validating dataset (step 2)

As it was necessary to choose training pixels objectively (without bias caused by a human factor) and in the most cases randomly with the possibility to repeat the selection (in the case of changes of the legend or different categorization of polygons into classes), a Python script was written also for this task (appendix II) using Python version 2.7 and ArcPy module. Based on the studied literature and the field mapping experience, 24 combinations of amount and sampling design of the training dataset were created (table 6). These combinations were tested in SW ENVI using the SVM classifier and its parameters which yielded the highest accuracy in the step 1 of the experiments. The combination with the highest overall accuracy was then used to classify the remaining locations (Husí boudy, Friesovy Boudy, Lahrovy boudy, Klínové Boudy and Zadní Rennerovky).

Table 6: The tested combinations of sampling design and amount of pixels (source: author).

	Random sampling design	Clustered sampling design	“Real life scenario” (RL)	Pixels at boundaries
50 pixels	Random 50 pixels	Clustered 50 pixels	RL 50 pixels	50 pixels at boundaries
100 pixels	Random 100 pixels	Clustered 100 pixels	RL 100 pixels	100 pixels at boundaries
200 pixels	Random 200 pixels	Clustered 200 pixels	RL 200 pixels	200 pixels at boundaries
2.5% of pixels	Random 2.5% of pixels	Clustered 2.5% of pixels	RL 2.5% of pixels	2.5% of pixels at boundaries
5% of pixels	Random 5% of pixels	Clustered 5% of pixels	RL 5% of pixels	5% of pixels at boundaries
10% of pixels	Random 10% of pixels	Clustered 10% of pixels	RL 10% of pixels	10% pixels at boundaries

5.2.1 The amount of training pixels

There are two sampling rates, which can be utilized when selecting a training dataset. One is called equal sample rate (ESR), in which a fixed percentage of pixels is sampled from

each class. And the other is called equal sample size (ESS), in which a fixed number of pixels are sampled from each class (Huang et al., 2002). Both of these strategies were utilized for the determination of the experimental training datasets in this master thesis. Furthermore another rule stated by Pal and Mather 2006 was followed. The relationship between the amount of training pixels and the amount of validating pixels is defined by this rule in a way, that there should be at least double amount of the validating pixels as there is training pixels (Pal and Mather, 2006). Thus the amount of the experimental training pixels was determined according to the references in the table 5, the stated rules and HW requirements as: 50, 100 and 200 pixels of each class and 2.5%, 5% and 10% of the pixels of each class (fig. 28). In the case there were not enough pixels in the class to select 50, 100 or 200 pixels and still have twice the amount left for the validation, 30% of the pixels of the particular class were chosen.

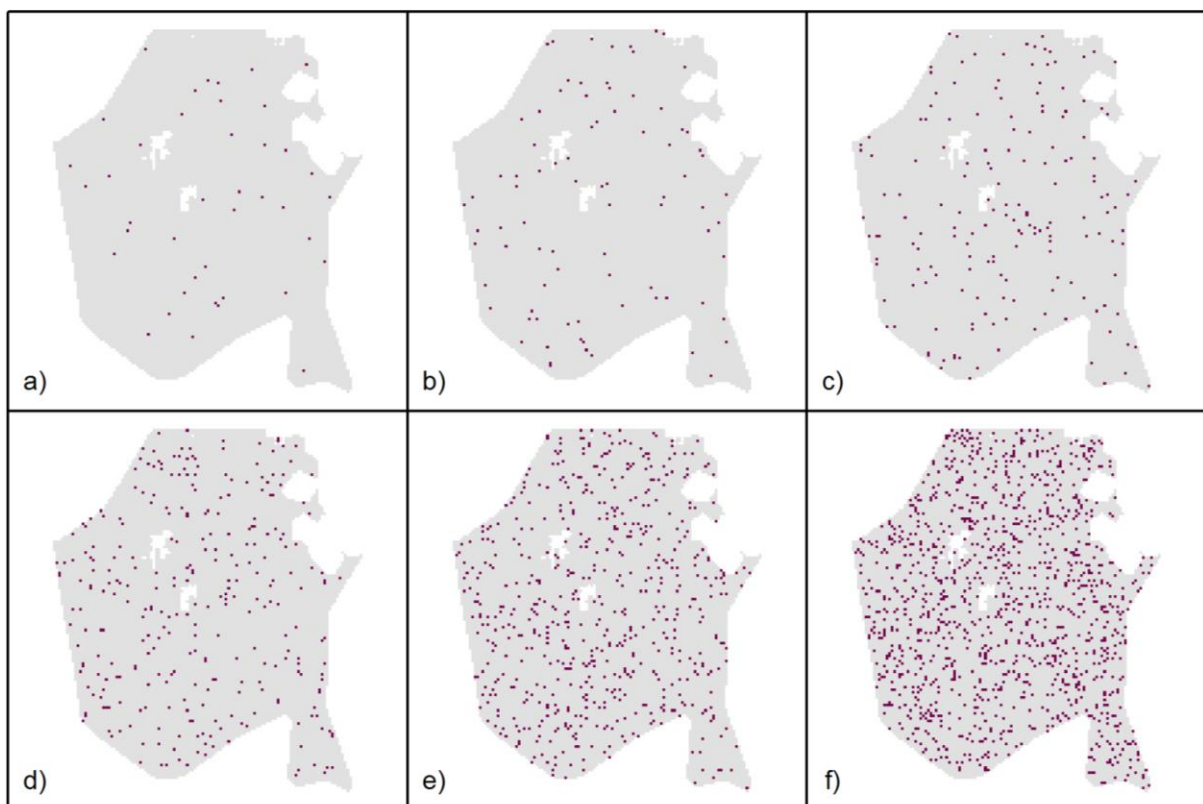


Fig. 28: The different amounts of pixels tested in the experiments shown in random stratified sampling design: a) 50 pixels, b) 100 pixels, c) 200 pixels, d) 2.5% of pixels in the class, e) 5% of pixels in the class and f) 10% of pixels in the class (source: own ArcGIS output).

5.2.2 Sampling design

There are several spatial sampling strategies, which are commonly used when determining training datasets. The most statistical tests generally assume a random spatial sampling, thus this type of sampling is considered as the most suitable. However randomly allocated points are often inaccessible because of difficult terrain or other obstacles, moreover small categories might be undersampled or missed. Another type of spatial sampling is clustered sampling. Clustered sampling can be a solution when there is limited time and resources for the field campaign as it reduces travel time in the field (Jones and Vaughan, 2010). Nevertheless the results can be spatially correlated when sampling sites are too close to each other (Jones and Vaughan, 2010; Huang et al., 2002). And like in the previous case, small categories can be missed (Jones and Vaughan, 2010). To prevent undersampling of small classes, we can choose stratified versions of sampling strategies (Jones and Vaughan, 2010; Belousov et al., 2002). This means that the study area is at first divided into polygons according to classes and then samples are collected according to chosen sampling design from each polygon. According to Jones and Vaughan (2010) stratified random sampling strategy is often the most efficient strategy.

In addition to the above mentioned strategies, also random boundary sampling (samples of mixed spectral responses selected from geographical boundaries of the classified communities) and “real life scenario” sampling (one or more bigger polygons grown from one or more seeding pixels to contain certain amount of pixels) were utilized in the experiments. The boundary sampling has been already used for classification of crops by the SVM classifier. Foody and Mathur (2004, 2006) tested the effect of border training patterns on the accuracy of SVM classification, using either geographical boundaries or boundaries defined by physical variables. It was learnt, that these purposefully selected samples can provide higher or comparable accuracy as conventionally defined training sets. The “real life scenario” spatial sampling was tested because it is convenient and common to collect training pixels in that way.

All the tested sampling strategies were performed in the stratified way.

5.2.2.1 Random sampling design

The randomly sampled pixels were determined using `random.sample` function. The IDs of the pixels were sampled from the list of IDs of all the pixels. These pixels were then

selected using `arcpy.SelectLayerByAttribute_management` and saved as a new vector layer (fig. 29). The selection was then switched and the rest of the pixels were used as validating pixels.

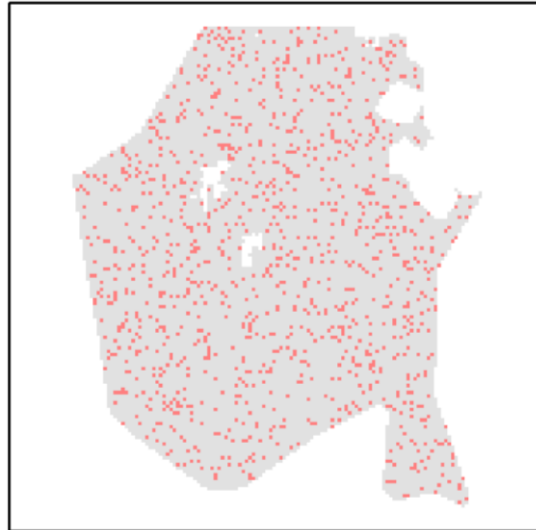


Fig. 29: Random sampling design (source: own ArcGIS output).

5.2.2.2 Clustered sampling design

The creation of clustered training pixels involved several steps (fig. 30). At first, the area from where the centres of cluster can be chosen was defined. This area contained every pixel which was further than 1 or 2 metres from the boundary of a polygon.

Secondly, the centres of clusters were randomly defined using the same method as is mentioned in chapter 5.2.2.1, only the amount of pixels was different (1% of the total amount of pixels in the class for ESR and 3, 6 and 12 centres for 50, 100 and 200 pixels, respectively). Then the neighbourhood around these centres was defined. 8-N neighbourhood was defined as all pixels within a distance of 0 metres from the centres and 24-N neighbourhood was defined as all pixels within a distance of 1 metre from the centres. The 8-N neighbourhood was utilized in the case of 2.5% and 5% of the total amount of pixels in the class and the 24-N neighbourhood was utilized in all the other cases.

Finally a certain amount of pixels was randomly selected from the defined neighbourhoods using the technique mentioned in chapter 5.2.2.1 and saved as training dataset. The remaining pixels were then used as a validating dataset.

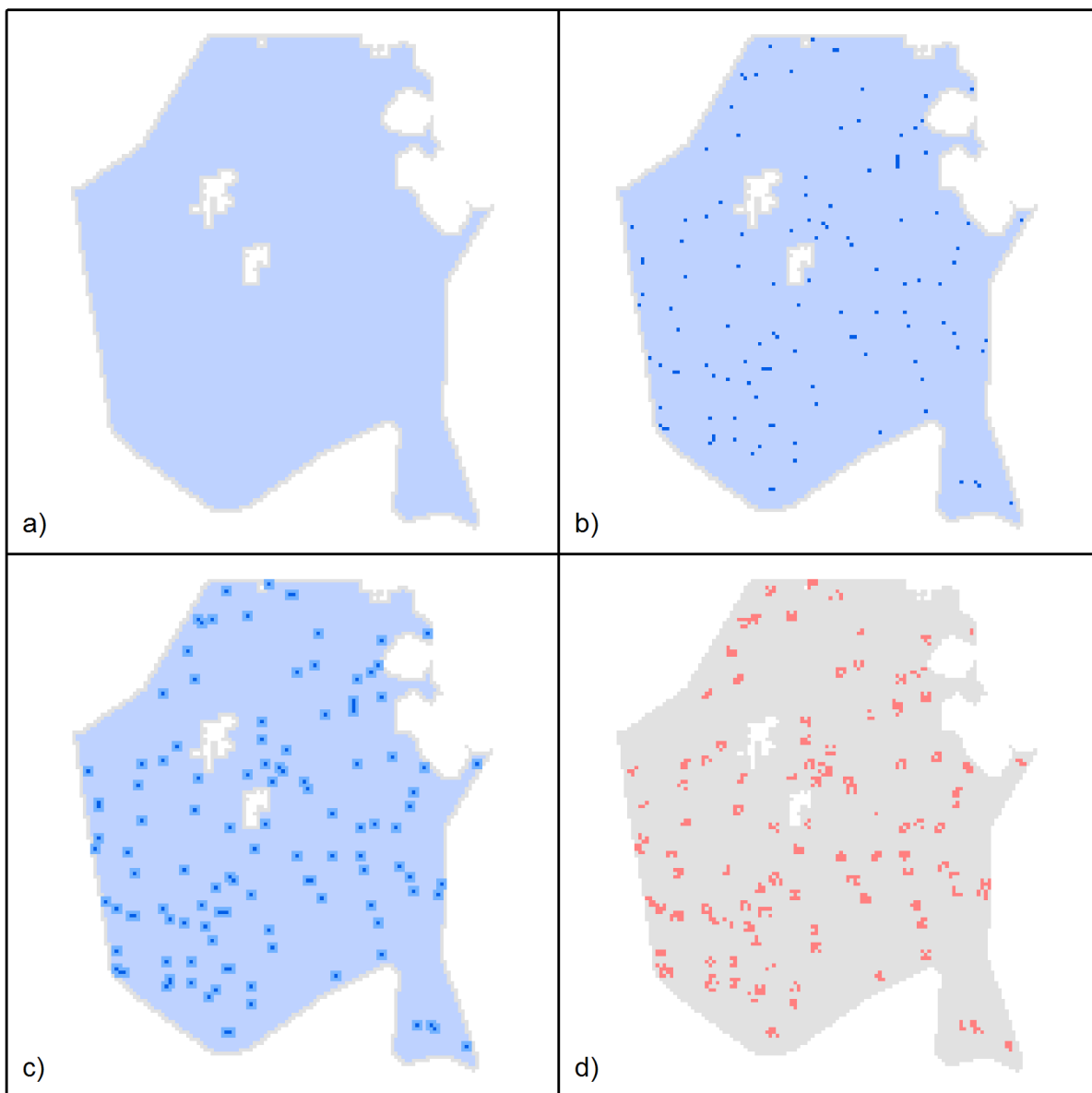


Fig. 30: Clustered sampling design: a) area from which centres of clusters can be chosen, b) randomly chosen centres of clusters c) neighbourhoods around randomly chosen centres, d) clusters of training pixels (source: own ArcGIS output).

5.2.2.3 “Real life scenario” sampling design

To simulate a common sampling strategy where polygons of spectrally uniform pixels are collected the “real life scenario” sampling design was tested (fig. 31). In the first place, one or more seeding pixels were manually chosen from the polygons, based on the notes of the botanist made during the field mapping. The seeding pixels were chosen in the polygons, which were supposed to be the most spectrally uniform and the amount of the seeding pixels depended on the number of sub-communities included in a class.

The polygons were then grown around these seeding pixels until they reached or exceeded the given amount of pixels. In majority, the given amount of pixels was exceeded so it was necessary to remove the exceeded number of pixels. This was done by defining outer line of a polygon, creating a list of IDs of the pixels contained in the line and then randomly selecting the desired amount of pixels, which were removed from the polygon. The pixels which left in the polygon were then used as the training dataset and all the other pixels as validating dataset.

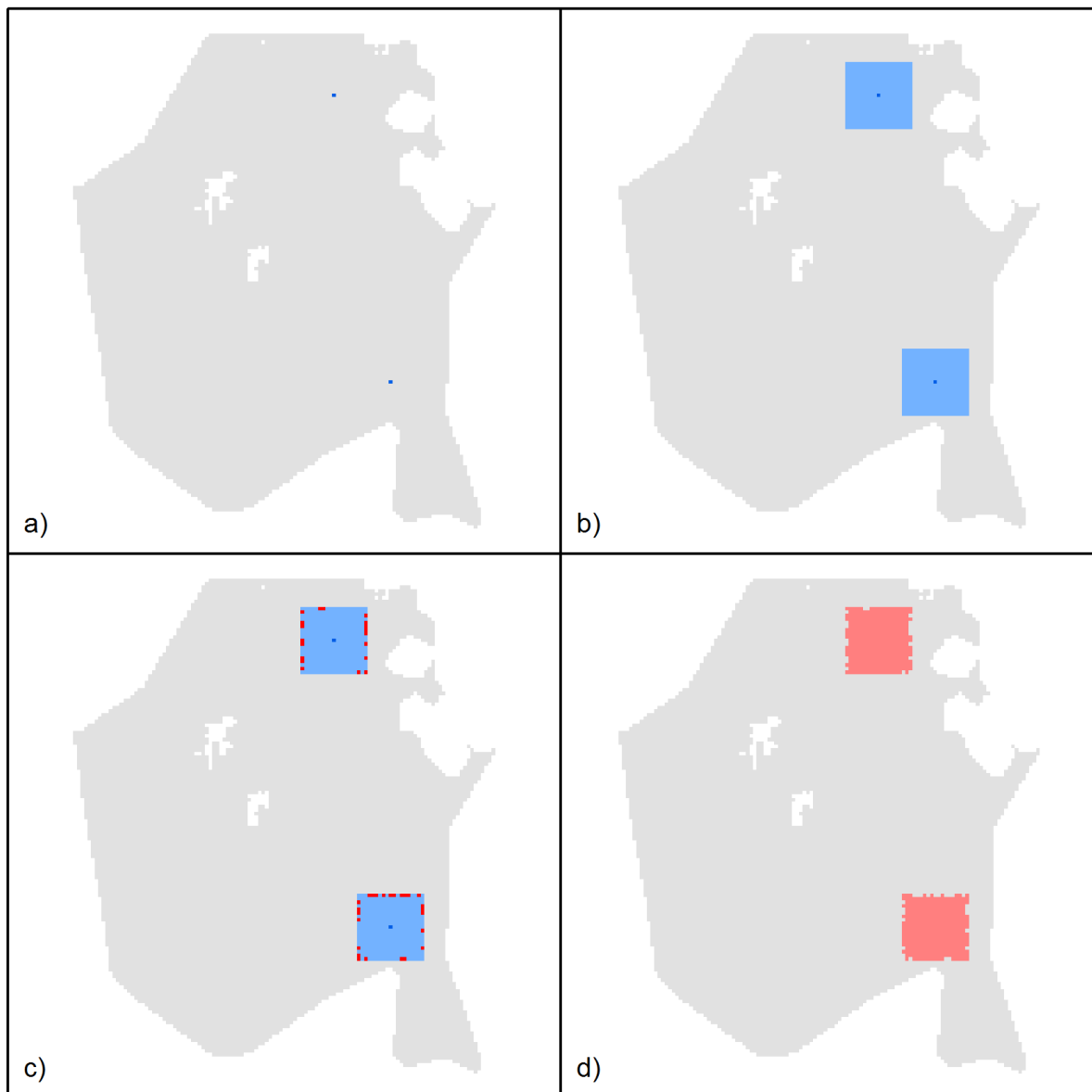


Fig. 31: “Real life scenario” sampling design: a) seeding pixels, b) grown polygons, c) grown polygons together with pixels selected to be removed (red), d) training dataset for “real life scenario” sampling design (source: own ArcGIS output).

5.2.2.4 Random pixels at boundaries

To create a training dataset with pixels at boundaries of classes (fig. 32), the boundary had to be identified first. All the pixels within 1 metre from the outer line of a polygon were considered as boundary pixels.

In the second step, a certain amount of pixels were randomly selected from the boundary pixels and the selection was saved as a new vector layer with training pixels. The rest of the pixels were utilized as validating pixels.

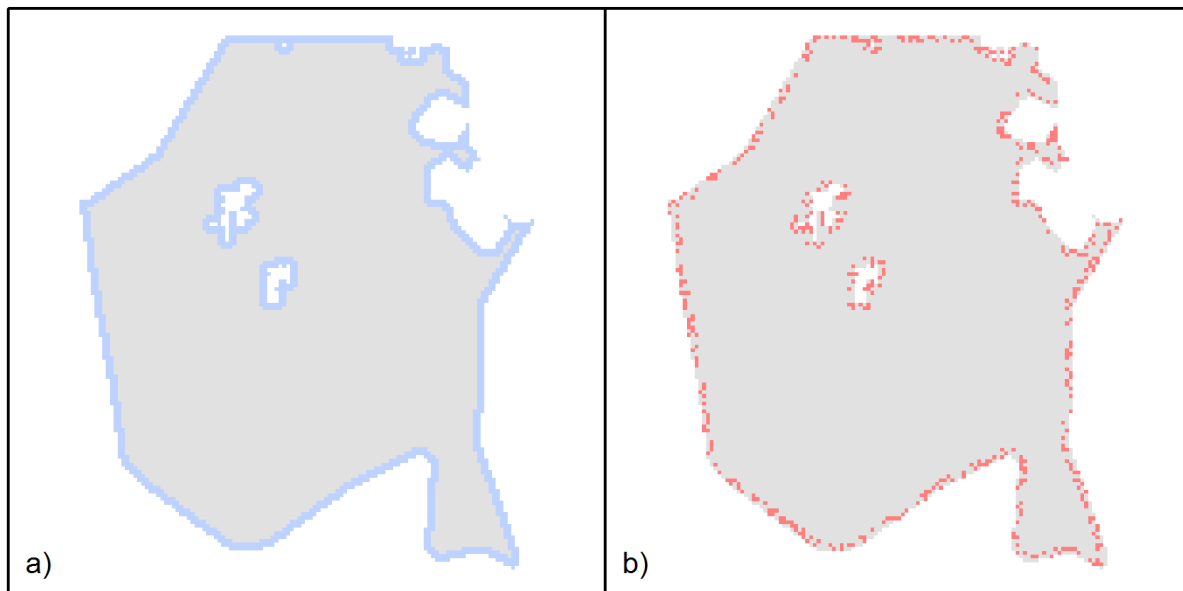


Fig. 32: Random pixels at boundaries: a) defined boundary from which training pixels are chosen randomly, b) training pixels at boundaries (source: own ArcGIS output).

5.3 ANN setup

The basic guidelines for setting parameters of NN classifier were adapted from Pomahačová 2012 , where it was learnt that the number of training iterations, training rate (TR) and training threshold contribution (TTC) have the highest impact on the accuracy of the classification of meadows in the Krkonoše mountains (Pomahačová, 2012). The training rate determines the magnitude of the adjustments of the weights. A higher rate speeds up the training, however it also increases the risk of oscillations or non-convergence of the training result. The training threshold contribution is used to adjust the changes to a node's internal weight. If TTC is set to 0, the algorithm does not adjust the node's internal weights.

Adjustments of the node's internal weight can lead to better classifications of a certain area of interest, but too many adjustments can also lead to a poor generalization ability of a classifier (ENVI, NN). The best values of these parameters were stated by Pomahačová as follows: 15000 training iterations, 0.1 for training rate and 0.2 for training threshold contribution (Pomahačová, 2012).

These values were used as initial values for testing the NN set up. However it was learnt that when 15000 iterations were used the computational time of the algorithm was too long with no increase in accuracy, probably due to high spatial resolution of the utilized data (1m) and many training pixels. Therefore only 1000 to 5000 iterations with an increment of 2000 were tested. The training rate of 0.1 and 0.2 was tested. The threshold training contribution was tested in a range from 0.2 to 0.8 with an increment of 0.2. Furthermore only one hidden layer was used as it is enough in the majority of cases and logistic activation function was utilized as it gave the best results in the experiments mentioned in the studied literature (chapter 2.3.4.1; Petropoulos et al., 2012; Camps-Valls et al., 2004). The rest of the parameters were left as default like in the case of Pomahačová (2012). The testing of NN setup was also performed on the enclave Přední Rennerovky and the best results were used to classify the remaining enclaves of the interest. The sampling design and amount of pixels which yielded the best classification accuracy for the SVM classifier was used also for the NN classifier.

5.4 Assessment of the accuracy of classification

The error matrix and kappa analysis were used to determine the accuracy of classifications. Moreover none of the training pixels was utilized as a reference pixel for the assessment of the accuracy of classification to avoid estimation of overoptimistic result of true classification accuracy (Jones and Vaughan, 2010).

The error matrix (also referred to as a confusion matrix or a classification matrix) contains several measures of accuracy. The first and the most straightforward measure is the overall accuracy (OA). OA is defined as the total amount of correctly classified pixels (values in the major diagonal) divided by the total amount of classified pixels. Furthermore also measures of accuracy for each class can be derived from the error matrix. Depending on what we are interested in, we can count the number of the correctly classified pixels of a particular class either as a fraction of the "true" number of pixels of that class, or as a fraction of the

number of the pixels classified in that class. The first option (correctly classified pixels divided by the “true” number of pixels of a particular class) is a measure of producer’s accuracy (PA). The complement of producer’s accuracy is called omission error (OE) and it indicates the number of pixels for a particular class which were misclassified into other classes. The PA together with omission error is usually noted in the columns at the far right of the error matrix. The second option (correctly classified pixels divided by the number of the pixels classified into a particular class) is a measure of user’s accuracy (UA) and it indicates the reliability of the classification of a particular class. The complement of user’s accuracy is called commission error (CE). The UA and commission error are usually noted in the lowermost rows of the error matrix (Jones and Vaughan, 2010).

The kappa analysis is a widely used method of accuracy assessment involving normalization of the data and taking into account omission and commission errors in addition to the values in the major diagonal. This is the main difference from the overall accuracy measurement. The kappa coefficient (κ) is calculated as:

$$kappa = \frac{n \sum_{i=1}^k x_{ii} - \sum_{i=1}^k (x_{i+} * x_{+i})}{n^2 - \sum_{i=1}^k (x_{i+} * x_{+i})}$$

where x_{ii} are the diagonal cells of the matrix, the x_{i+} are the row marginal totals and the x_{+i} are the column marginal totals and the n is the total number of elements in the matrix. The kappa coefficient is a measure of agreement after chance agreement is removed. Therefore kappa is used to determine if the classification is significantly better than chance. Kappa coefficient greater than 0.75 means that classifier performs well, on the other hand kappa coefficient lower than 0.40 suggests poor performance of a classifier (Jones and Vaughan, 2010).

6. Results

6.1 The SVM classifier

6.1.1. The parameters of the SVM classifier

The results of SVM parameters testing on the model enclave of Přední Rennerovky are summarized in the tables 7.A, 7.B and 7.C. With regard to given conditions (amount of training pixels and their sampling design), it can be said that the worst performing kernel was linear kernel (the highest achieved OA 69.44 % and kappa 0.58; table 7.C). The performance of the RBF kernel was mediocre (the highest achieved OA 71.11% and kappa 0.60; table 7.B) and the best performing kernel was polynomial kernel of polynomial degree 6 (the highest achieved OA 74.46% and kappa 0.64; table 7.A).

As for the influence of changes of gamma parameter on accuracy of classification, the result was the same for both tested kernels (RBF and polynomial). The changes did not have any effect on the accuracy of classification, even if the whole scale of values available in ENVI was tested.

Penalty parameter and the polynomial degree of polynomial kernel had, on the other hand, major effect on the accuracy of classification. The classification was more accurate with the higher polynomial degree of polynomial kernel used (for a comparison: OA 69.63% and kappa 0.58 when polynomial degree of 2 was used and OA 73.71% and kappa 0.63 when polynomial degree of 6 was used). The trend of penalty parameter was up to few exceptions similar to the one of polynomial degree – the higher the value of penalty parameter, the higher the accuracy of the classification. The penalty parameter of value 100 yielded those results for RBF and polynomial kernels mentioned as the best in the beginning of this chapter. The only exception was linear kernel, where the accuracy of classification was varying as the value of penalty parameter increased. However the variation was insignificant, only on the scale of decimal digits. Thus it could be concluded, that the changes in the values of penalty parameter did not affect the accuracy of the classification when linear kernel was utilized.

The polynomial kernel of the 6th polynomial degree with default value of gamma parameter and value of 100 of penalty parameter was used for further experiments and classifications.

Table 7.A. The results of accuracy assessment for tested combinations of parameters of polynomial kernel. The value of 0.143 for gamma parameter is the default value set by ENVI 4.7 (1/7 bands) (source: own ENVI analysis).

# of case	Polynomial degree	Gamma parameter	Penalty parameter	OA (%)	Kappa
1.	1	0.143	50	68.97	0.57
2.	2	0.143	50	69.63	0.58
3.	3	0.143	50	71.17	0.60
4.	4	0.143	50	72.32	0.61
5.	5	0.143	50	73.52	0.63
6.	6	0.143	50	73.71	0.63
7.	6	0.010	50	73.71	0.63
8.	6	0.100	50	73.71	0.63
9.	6	1.000	50	73.71	0.63
10.	6	10.000	50	73.71	0.63
11.	6	100.000	50	73.71	0.63
12.	6	1000.000	50	73.71	0.63
13.	6	10000.000	50	73.71	0.63
14.	6	0.143	0 (0.01)	33.04	0.25
15.	6	0.143	10	71.72	0.61
16.	6	0.143	20	73.12	0.62
17.	6	0.143	30	73.51	0.63
18.	6	0.143	40	73.60	0.63
19.	6	0.143	60	73.56	0.63
20.	6	0.143	70	73.78	0.63
21.	6	0.143	80	74.00	0.63
22.	6	0.143	90	74.34	0.64
23.	6	0.143	100	74.46	0.64

Table 7.B. The results of accuracy assessment for tested combinations of parameters of RBF kernel. The value of 0.143 for gamma parameter is the default value set by ENVI 4.7 (1/7 bands) (source: own ENVI analysis).

# of case	Gamma parameter	Penalty parameter	OA (%)	Kappa
1.	0.010	50	70.01	0.59
2.	0.100	50	70.01	0.59
3.	1.000	50	70.01	0.59
4.	10.000	50	70.01	0.59
5.	100.000	50	70.01	0.59
6.	1000.000	50	70.01	0.59
7.	10000.000	50	70.01	0.59
8.	0.143	0 (0.01)	54.51	0.00
9.	0.143	10	68.97	0.57
10.	0.143	20	69.43	0.59
11.	0.143	30	69.73	0.58
12.	0.143	40	69.97	0.58
13.	0.143	50	70.00	0.59
14.	0.143	60	70.23	0.59
15.	0.143	70	70.37	0.59
16.	0.143	80	70.61	0.59
17.	0.143	90	70.82	0.59
18.	0.143	100	71.11	0.60

Table 7.C. The results of accuracy assessment for linear kernel and different values of penalty parameter (source: own ENVI analysis).

# of case	Penalty parameter	OA (%)	Kappa
1.	0 (0.01)	22.31	0.02
2.	10	68.96	0.57
3.	20	69.44	0.58
4.	30	69.33	0.58
5.	40	69.35	0.58
6.	50	69.40	0.58
7.	60	69.35	0.58
8.	70	69.42	0.58
9.	80	69.39	0.58
10.	90	69.29	0.58
11.	100	69.24	0.58

6.1.2 Sampling design and the amount of training pixels

The overall accuracies and kappa coefficients for different combinations tested are noted in the table 8. Regarding various sampling designs, the random sampling design gave the best results (given a certain amount of pixels) in most of the cases (OA 83.63% and kappa 0.74 for 10% of pixels). However also the clustered sampling design and “pixels at boundaries” sampling design gave satisfactory results (OA 82.11 % and kappa 0.71 and OA 81.68% and kappa 0.70 respectively for 10% of pixels). The worst accuracies were measured for the “real life scenario” sampling design (OA 74.48% and kappa 0.60 for 10% of pixels). The only exception for the “real life scenario” occurred when only 50 pixels were used as a training sample. In this situation the accuracy achieved by utilizing this sampling design was the higher than the ones achieved by other sampling designs. The combination of the “real life scenario” sampling design and 50 pixels as a training sample was exceptional also with regard to amount of 50 pixels, because this situation was also the case when the highest accuracy was measured for this amount of training pixels. Otherwise the greater amount of training pixels resulted in higher accuracies.

The combinations, which yielded the best results were as follows: randomly sampled 2.5%, 5% and 10% of pixels (OA close to and higher than 82% and kappa 0.70, 0.73 and 0.74 respectively), clustered 2.5%, 5% and 10% of pixels (OA around 82% and kappa 0.71, 0.72 and 0.71 respectively) and 2.5%, 5% and 10% of pixels at boundaries (OA over 80% and kappa 0.67, 0.67 and 0.70 respectively). Randomly sampled 10% of pixels combination was used to classify the remaining enclaves, because of the highest OA reached.

In general training pixels selected by ESR method gave higher accuracies than training pixels selected by ESS method.

Table 8. The results of accuracy assessment for tested combinations of sampling designs and the amount of pixels noted in table 6 (source: own ENVI analysis).

Combination	OA (%)	Kappa	Combination	OA (%)	Kappa
Random 50 pixels	69.32	0.58	RL 50 pixels	75.23	0.62
Random 100 pixels	73.36	0.63	RL 100 pixels	71.25	0.58
Random 200 pixels	74.46	0.64	RL 200 pixels	65.87	0.53
Random 2.5% of pixels	81.80	0.70	RL 2.5% of pixels	73.53	0.59
Random 5% of pixels	83.17	0.73	RL 5% of pixels	74.17	0.59
Random 10% of pixels	83.63	0.74	RL 10% of pixels	74.48	0.60
Clustered 50 pixels	58.19	0.44	50 pixels at boundaries	68.13	0.56
Clustered 100 pixels	60.93	0.47	100 pixels at boundaries	71.00	0.59
Clustered 200 pixels	71.11	0.52	200 pixels at boundaries	73.11	0.62
Clustered 2.5% of pixels	81.80	0.71	2.5% of pixels at boundaries	80.25	0.67
Clustered 5% of pixels	82.42	0.72	5% of pixels at boundaries	80.47	0.67
Clustered 10% of pixels	82.11	0.71	10% of pixels at boundaries	81.68	0.70

6.1.3 The accuracy assessment of the classification of the enclaves of interest

The most important characteristics of the accuracy of the classification of the areas of interest can be found in tables 9 to 14. The more detailed characteristics are in the appendices III to XIV. Meaning of the abbreviations used in the tables is as follows: OA – overall accuracy, PA – producer’s accuracy, UA – user’s accuracy, CE – commission error, OE – omission error.

The highest OA and kappa coefficients were measured for the model location (Přední Rennerovky) and for the location of Friesovy boudy, which were re-checked and training polygons for these locations were fixed in the best possible way given material resources. The OA for these locations were 83.63% and 87.63% respectively. Kappa coefficients were 0.74 and 0.75 respectively. Mediocre results were achieved on locations Husí Boudy (OA 61.19% and kappa 0.48), Lahrový Boudy (OA on both flight lines around 65.40 % and kappa on both flight lines around 0.50) and for the part of Zadní Rennerovky on flight line 10_20 (OA 69.67% and kappa 0.53). The worst accuracies were obtained for location of Klínové Boudy (OA 68.51% but kappa only 0.33) and for Zadní Rennerovky on flight line 10_58 (OA

59.34% and kappa 0.32). Generally, smaller locations were classified with the higher accuracy than bigger locations.

Table 9. The results of accuracy assessment for the location of Přední Rennerovky (SVM classifier). Numbers of classes: 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species. **OA = 83.63%** and **kappa = 0.74** (source: own ENVI analysis).

	UA (%)	CE (%)	PA (%)	OE (%)
2	81.22	18.78	80.03	19.97
3	90.22	9.78	85.90	14.10
4	71.70	28.30	45.05	54.95
5	78.50	21.50	55.93	44.07
6	83.39	16.61	93.26	6.74

Table 10. The results of accuracy assessment for the location of Friesovy Boudy (SVM classifier). Numbers of classes: 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species. **OA = 87.63%** and **kappa = 0.75** (source: own ENVI analysis).

	UA (%)	CE (%)	PA (%)	OE (%)
4	0.00	0.00	0.00	100
5	88.91	11.09	83.50	16.50
6	86.76	13.24	92.07	7.93
7	0.00	0.00	0.00	100

Table 11. The results of accuracy assessment for the location of Husí Boudy (SVM classifier). Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species. **OA = 61.19%** and **kappa = 0.48** (source: own ENVI analysis).

	UA (%)	CE (%)	PA (%)	OE (%)
1	69.86	30.14	66.38	33.62
2	56.78	43.22	73.44	26.56
3	63.97	36.03	71.53	28.47
4	51.17	48.83	31.02	69.98
5	40.76	59.24	12.07	87.93
6	50.00	50.00	7.82	92.18

Table 12. The results of accuracy assessment for the location of Klínové Boudy (SVM classifier). Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species; 8 – vegetation of springs; 9 – subalpine tall-fern vegetation. **OA = 68.51%** and **kappa = 0.33** (source: own ENVI analysis).

	UA (%)	CE (%)	PA (%)	OE (%)
1	0.00	0.00	0.00	100.00
2	59.27	40.73	16.85	83.15
4	59.67	40.33	20.03	79.97
5	62.80	37.20	44.95	55.05
6	70.43	29.57	92.20	7.80
7	55.22	44.78	13.09	86.91
8	42.18	57.82	11.77	88.23
9	100.00	0.00	0.43	99.57

Table 13. The results of accuracy assessment for the location of Lahrový Boudy (SVM classifier). Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species; 8 – vegetation of springs. **OA_{line 08_43} = 65.33%** and **kappa_{line 08_43} = 0.44**. **OA_{line 10_45} = 65.45%** and **kappa_{line 10_45} = 0.50** (source: own ENVI analysis).

Line 08_43	UA (%)	CE (%)	PA (%)	OE (%)
1	55.14	44.86	41.62	58.38
2	69.26	30.74	21.65	78.35
3	73.77	26.23	79.74	20.26
4	51.80	48.20	10.55	89.45
5	53.63	46.37	65.37	34.63
6	72.68	27.32	34.15	65.85
7	0.00	0.00	0.00	100.00
Line 10_45	UA (%)	CE (%)	PA (%)	OE (%)
1	72.16	27.84	84.25	15.75
2	50.82	49.18	31.60	68.40
3	68.50	31.50	86.01	13.99
4	44.21	55.79	2.96	97.04
5	61.35	38.65	46.60	53.40
6	51.19	48.81	38.04	61.96
7	0.00	0.00	0.00	100.00
8	0.00	0.00	0.00	100.00

Table 14. The results of accuracy assessment for the location of Zadní Rennerovky (SVM classifier). Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species; 8 – vegetation of springs; 9 – subalpine tall-fern vegetation. **OA_{line 10_58} = 59.34%** and **kappa_{line 10_58} = 0.32**. **OA_{line 10_20} = 69.67%** and **kappa_{line 10_20} = 0.53** (source: own ENVI analysis).

Line 10_58	UA (%)	CE (%)	PA (%)	OE (%)
1	69.38	30.62	49.41	50.59
2	82.95	17.05	66.17	33.83
3	68.01	31.99	16.54	83.46
4	59.78	40.22	18.37	81.63
5	53.68	46.32	12.39	87.61
6	57.03	42.97	93.17	6.83
7	73.44	26.56	38.07	61.93
8	63.36	36.64	19.19	80.81
9	21.43	78.57	0.53	99.47
Line 10_20	UA (%)	CE (%)	PA (%)	OE (%)
1	83.63	16.37	79.01	20.99
2	70.91	29.09	57.38	42.62
3	63.03	36.97	32.18	67.82
4	25.93	74.07	1.51	98.49
5	63.97	36.03	16.11	83.89
6	61.67	38.33	86.72	13.28
7	43.37	56.63	10.07	89.93

The best classified category on the majority of locations was class 6 – degraded meadows with dominating grass species (e.g. accuracy 93.26% for Přední Rennerovky, 92.07% for Friesovy Boudy or 92.20% for Klínové Boudy). Only on locations Husí Boudy and Lahrovy Boudy (both flight lines), class 2 – oligotrophic grasslands (Husí Boudy, accuracy 73.44%) and class 3 – mesophile grasslands (Lahrovy Boudy on flight line 08_43 with accuracy 79.74% and Lahrovy Boudy on flight line 10_45 with accuracy 86.01%) were mapped the most accurately. The less abundant classes were usually classified with the lowest accuracy, such as class 7 (stands dominated by *Vaccinium* species), class 8 (springs) and class 9 (tall-fern vegetation). These classes were mapped with accuracies from 0% to 13.09%. In approximately half of the cases also waterlogged grasslands (class 4) and in some cases also degraded meadows dominated by *Dicotyledons* (class 5) were classified with the accuracy lower than 10%, because pixels belonging to these classes were confused the most between each other. Also pixels of class 3 were quite often confused with pixels of class 5. Other classes, which were occasionally confused amongst themselves, were classes 1, 2 and 6.

As for differences of results for enclaves on two flight lines (Lahrovy Boudy and Zadní Rennerovky), they vary greatly between enclaves. The both flight lines of Lahrovy Boudy were classified with almost the same accuracies ($\Delta OA = 0.11\%$ and $\Delta kappa = 0.05$), while there are discrepancies between flight lines of Zadní Rennerovky ($\Delta OA = 10.33\%$ and $\Delta kappa = 0.21$). All the classes, except class 1, on both lines of Lahrovy Boudy were mapped with quite a similar accuracy, whereas on location of Zadní Rennerovky it was only classes 2, 5 and 6 which were mapped with similar accuracy.

6.2 The ANN classifier

6.2.1 The parameters of the ANN classifier

The accuracies of classification for different combinations of NN classifier parameters can be seen in table 15.

As it was mentioned before, it was learnt during the testing that while number of iterations greater than 3000 do not significantly increase the accuracy of classification ($\Delta OA = 0.26\%$ and $\Delta kappa = 0.01$), it significantly increases the computational time necessary for the classification. Therefore the 3000 iterations were established as the most suitable number of iterations. Furthermore increasing of training rate caused a major drop in the accuracy of

classification, which supports the findings of Pomahačová (2012) and thus the TR of 0.1 like in the case of Pomahačová was utilized. From the tested possibilities of TTC, the value of 0.4 gave the best results and therefore other classifications were performed using TTC of value 0.4. The resultant combination of parameters of NN classifier was then as follows: 3000 iterations, 0.1 for TR and 0.4 for TTC with all the other parameters left as default.

Table 15. The results of accuracy assessment for tested combinations of parameters of the NN classifier (source: own ENVI analysis).

# of case	# of iterations	TR	TTC	OA (%)	Kappa
1.	1000	0.1	0.2	78.23	0.64
2.	3000	0.1	0.2	80.20	0.68
3.	5000	0.1	0.2	80.46	0.69
4.	3000	0.2	0.2	77.66	0.63
5.	3000	0.1	0.4	80.40	0.68
6.	3000	0.1	0.6	80.00	0.66
7.	3000	0.1	0.8	80.01	0.67

6.2.2 The accuracy assessment of the classification of the enclaves of interest

Some of the characteristics of the accuracy assessment of the areas of interest can be found in tables 16 to 21. The more detailed characteristics can be found in the appendices III to XIV. Meaning of the abbreviations used in the tables is as follows: OA – overall accuracy, PA – producer’s accuracy, UA – user’s accuracy, CE – commission error, OE – omission error.

The highest accuracies were obtained, like in the case of SVM classifier, for the model location of Přední Rennerovky (OA 80.37% and kappa 0.68) and for Friesovy Boudy (OA 87.35% and kappa 0.75). Average results were achieved for enclaves Husí Boudy (OA 58.86% and kappa 0.44), Lahrový Boudy on flight line 10_45 (OA 61.48% and kappa 0.45) and Zadní Rennerovky on flight line 10_20 (OA 67.61% and kappa 0.49). The average to worse results was gained for Lahrový Boudy on flight line 08_43 (OA 60.81% and kappa 0.39). The lowest values of OA and kappa coefficients were measured for the remaining enclaves: Zadní Rennerovky on flight line 10_58 (OA 57.86% and kappa 0.28) and Klínové Boudy (OA 66.20% but kappa 0.31). Again, it could be stated that in majority smaller locations were classified with the higher accuracy than bigger locations.

The class with the best classification results on most of the enclaves was also class 6 (e.g. accuracy 92.24% for Přední Rennerovky, or 95.17% for Zadní Rennerovky on flight line 10_58) and also the locations Husí Boudy and both flight lines of Lahrový Boudy were

exceptions regarding this fact. However different classes were mapped the most accurately on these enclaves than in the case of SVM classifier (class 3 – mesophile grasslands on location Husí Boudy with accuracy of 82.45%, class 5 – degraded meadows dominated by *Dicotyledons* on flight line 08_43 of Lahrový Boudy with accuracy of 90.15% and class 1 – *N.stricta* grasslands on flight line 10_45 of Lahrový Boudy with accuracy of 79.24%). Besides the already mentioned, less abundant classes, which were classified with lower accuracy by SVM classifier (classes 7 – stands dominated by *Vaccinium* species, 8 – springs and 9 – tall-fern vegetation), also class 4 (waterlogged grasslands) was often classified incorrectly. The mapping accuracies of these categories were usually between 0 to 11.96%. Occasionally also class 5 (Husí Boudy and Zadní Rennerovky on flight line 10_20) or classes 1 or 2 (Klínové Boudy and Lahrový Boudy) were classified with the accuracies close to or under 10%. The reason for this is the same as in the case of SVM classifier and that is the confusion of the pixels belonging to these classes with pixels belonging to classes with similar spectral characteristics.

Table 16. The results of accuracy assessment for the location of Přední Rennerovky (NN classifier). Numbers of classes: 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species. **OA = 80.37%** and **kappa = 0.68** (source: own ENVI analysis).

	UA (%)	CE (%)	PA (%)	OE (%)
2	67.06	32.94	80.74	19.26
3	95.74	4.26	73.92	26.08
4	70.56	29.44	37.57	62.43
5	70.42	29.58	56.21	43.79
6	81.00	19.00	92.24	7.76

Table 17. The results of accuracy assessment for the location of Friesový Boudy (NN classifier). Numbers of classes: 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species. **OA = 87.35%** and **kappa = 0.75** (source: own ENVI analysis).

	UA (%)	CE (%)	PA (%)	OE (%)
4	0.00	0.00	0.00	100.00
5	85.23	14.77	87.58	12.42
6	89.05	10.95	88.41	11.59
7	0.00	0.00	0.00	100.00

Table 18. The results of accuracy assessment for the location of Husí Boudy (NN classifier). Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species. **OA = 58.90%** and **kappa = 0.44** (source: own ENVI analysis).

	UA (%)	CE (%)	PA (%)	OE (%)
1	73.39	26.61	65.47	34.53
2	56.22	43.78	66.18	33.82
3	59.24	40.76	82.45	17.55
4	26.69	73.31	11.96	88.04
5	13.68	86.32	0.43	99.57
6	6.44	93.55	2.23	97.77

Table 19. The results of accuracy assessment for the location of Klínové Boudy (NN classifier). Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species; 8 – vegetation of springs; 9 – subalpine tall-fern vegetation. **OA = 66.20%** and **kappa = 0.31** (source: own ENVI analysis).

	UA (%)	CE (%)	PA (%)	OE (%)
1	0.00	0.00	0.00	100.00
2	0.00	0.00	0.00	100.00
4	47.34	52.66	31.84	68.16
5	48.78	51.22	48.70	51.30
6	71.77	28.23	88.30	11.70
7	41.30	58.70	0.14	99.86
8	22.19	77.81	0.46	99.54
9	0.00	0.00	0.00	100.00

Table 20. The results of accuracy assessment for the location of Lahrovy Boudy (NN classifier). Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species; 8 – vegetation of springs. **OA_{line 08_43} = 60.81%** and **kappa_{line 08_43} = 0.39**. **OA_{line 10_45} = 61.48%** and **kappa_{line 10_45} = 0.45** (source: own ENVI analysis).

Line 08_43	UA (%)	CE (%)	PA (%)	OE (%)
1	53.21	46.79	31.17	68.83
2	72.30	27.70	10.40	89.60
3	85.11	14.89	59.31	40.69
4	0.00	0.00	0.00	100.00
5	46.35	53.65	90.15	9.85
6	89.57	10.43	21.22	78.78
7	0.00	0.00	0.00	100.00
Line 10_45	UA (%)	CE (%)	PA (%)	OE (%)
1	66.61	33.39	79.24	20.76
2	44.45	55.55	11.89	88.11
3	73.73	26.27	75.61	24.39
4	0.00	0.00	0.00	100.00
5	48.12	51.88	56.78	43.22
6	41.68	58.32	38.52	61.48
7	0.00	0.00	0.00	100.00
8	0.00	0.00	0.00	100.00

Table 21. The results of accuracy assessment for the location of Zadní Rennerovky (NN classifier). Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species; 8 – vegetation of springs; 9 – subalpine tall-fern vegetation. $OA_{\text{line 10}_58} = 57.86\%$ and $kappa_{\text{line 10}_58} = 0.28$. $OA_{\text{line 10}_20} = 67.60\%$ and $kappa_{\text{line 10}_20} = 0.49$ (source: own ENVI analysis).

Line 10_58	UA (%)	CE (%)	PA (%)	OE (%)
1	72.13	27.87	43.14	56.86
2	77.63	22.37	12.49	87.51
3	0.00	0.00	0.00	100.00
4	52.01	47.99	9.05	90.95
5	79.85	20.15	65.20	34.80
6	55.51	44.49	95.17	4.83
7	62.56	37.44	35.10	64.90
8	0.00	0.00	0.00	100.00
9	0.00	0.00	0.00	100.00
Line 10_20	UA (%)	CE (%)	PA (%)	OE (%)
1	80.67	19.33	79.59	20.41
2	64.29	35.71	46.83	53.17
3	63.59	36.41	22.36	77.64
4	0.00	0.00	0.00	100.00
5	46.32	53.68	5.61	94.39
6	59.92	40.08	86.60	13.40
7	27.79	72.21	2.71	97.29

A rather similar trend in the differences between flight lines of the same locations was observed also in the case of NN classifier with the discrepancies between flight lines of Zadní Rennerovky greater than those between flight lines of Lahrový Boudy ($\Delta OA = 9.75\%$ and $\Delta kappa = 0.2$ for Zadní Rennerovky versus $\Delta OA = 0.67\%$ and $\Delta kappa = 0.06$ for Lahrový Boudy).

6.3 Classification maps for the selected cases

Classification maps for the best combination for each SVM kernel type tested and also the classification maps for the best result for each sampling design are shown in this chapter. Further classification maps for all enclaves are shown in appendices.

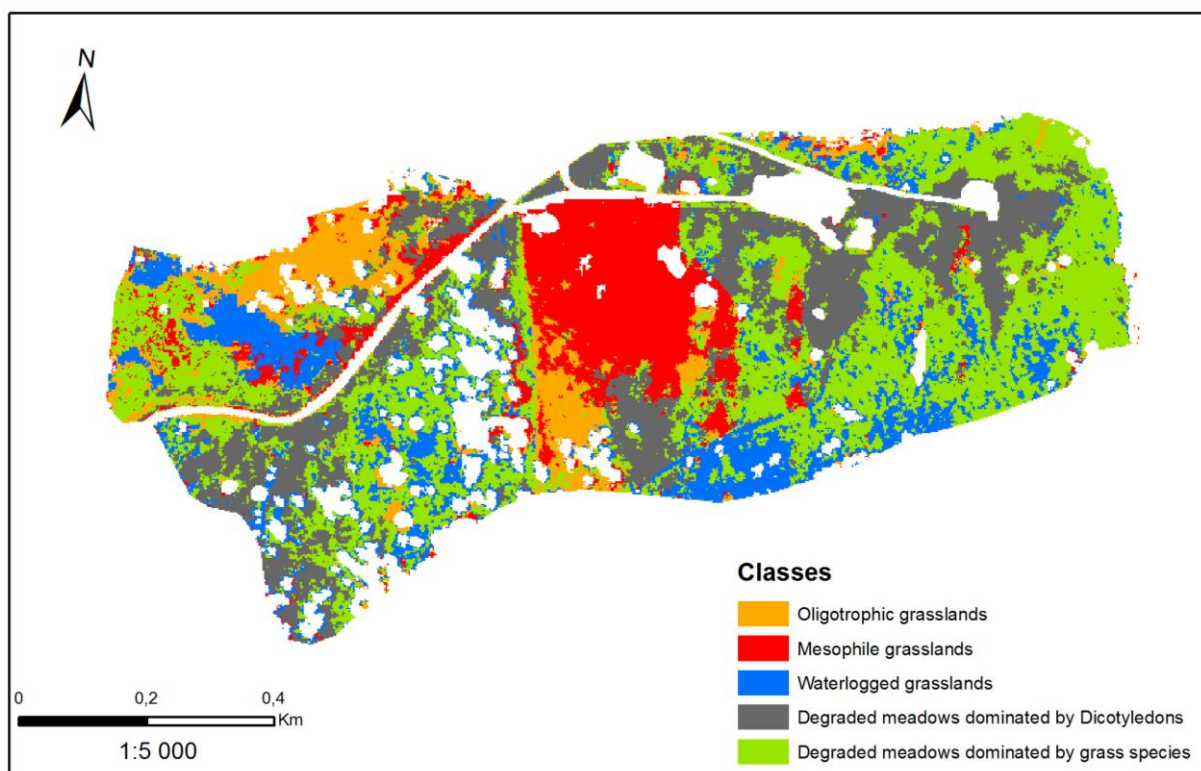
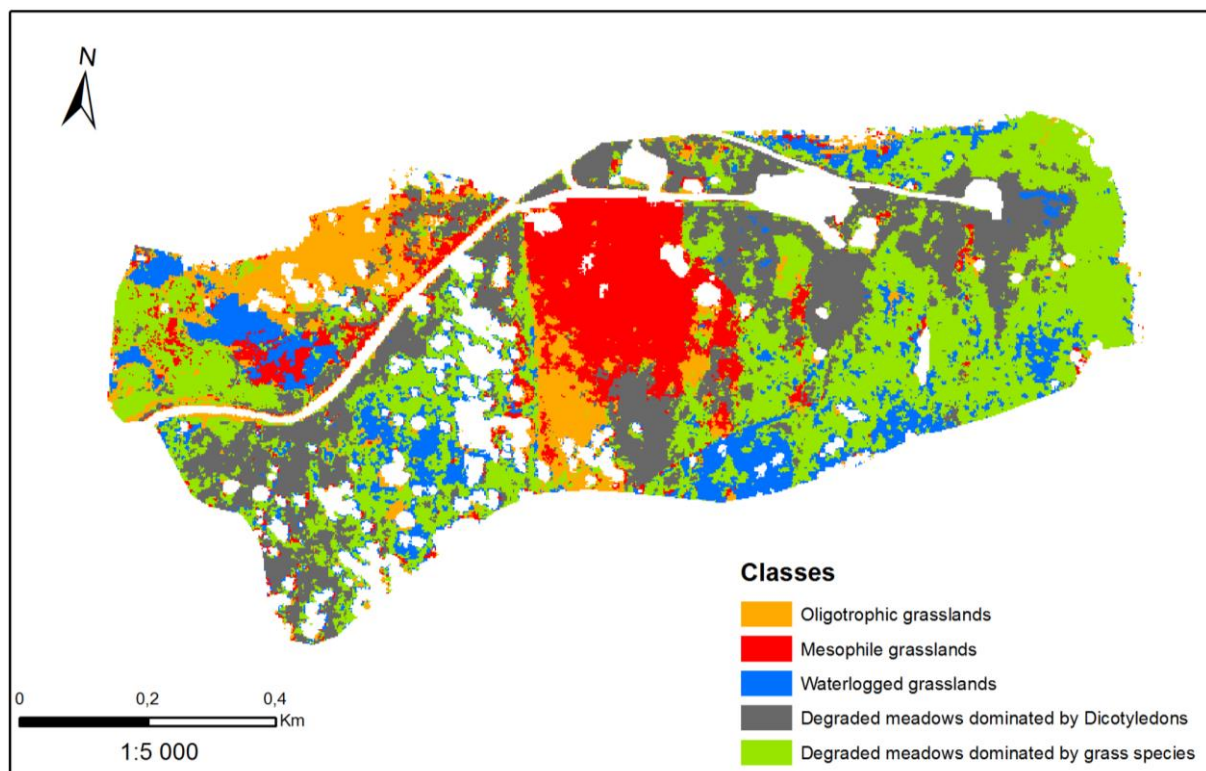


Fig.33: Classification maps for Přední Rennerovky for best combination of parameters of SVM polynomial kernel (upper) and best combination of parameters of SVM radial basis function kernel (lower).

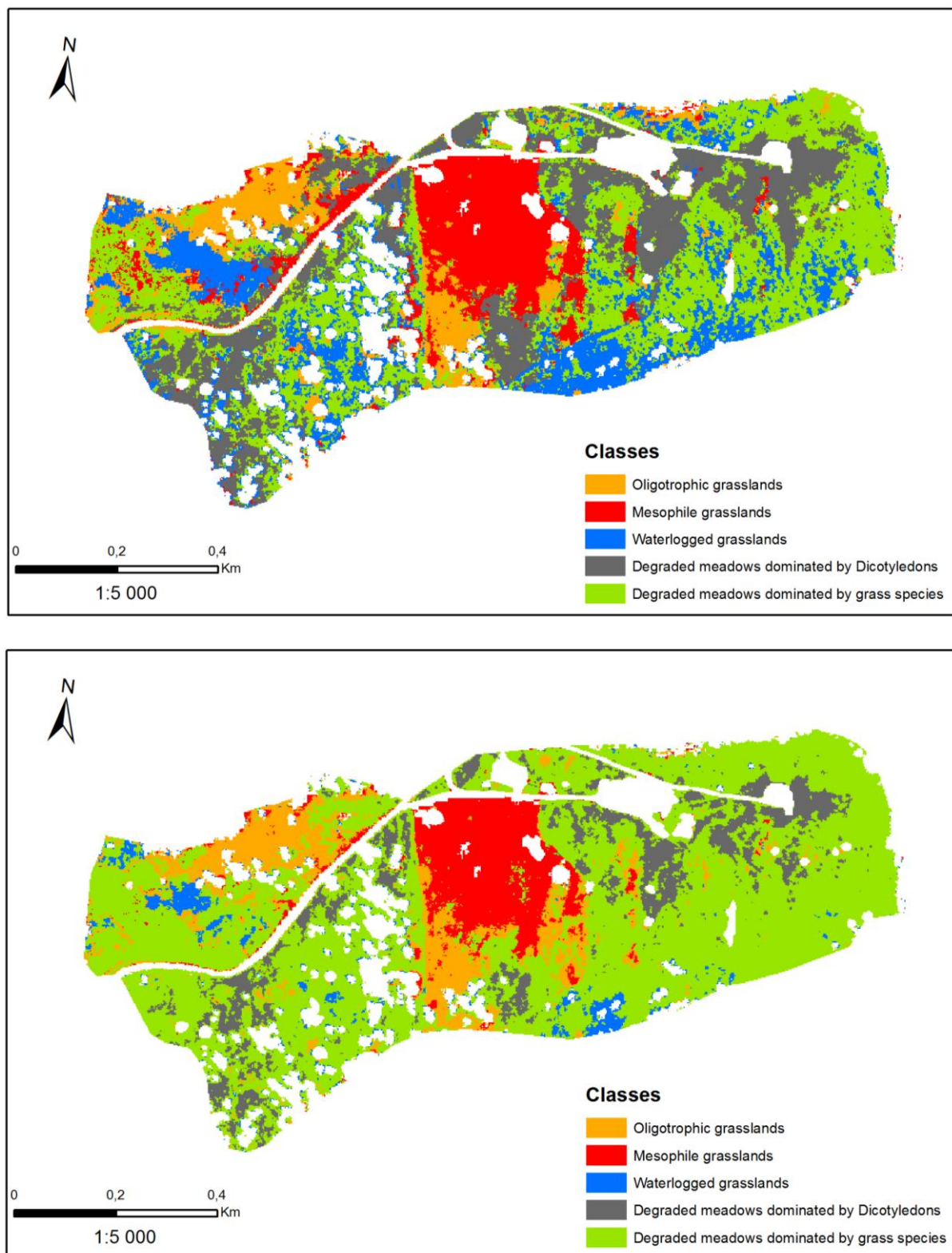


Fig. 34: Classification maps for Přední Rennerovky for best combination of parameters of SVM linear kernel (upper) and best combination of NN classifier parameters (lower).

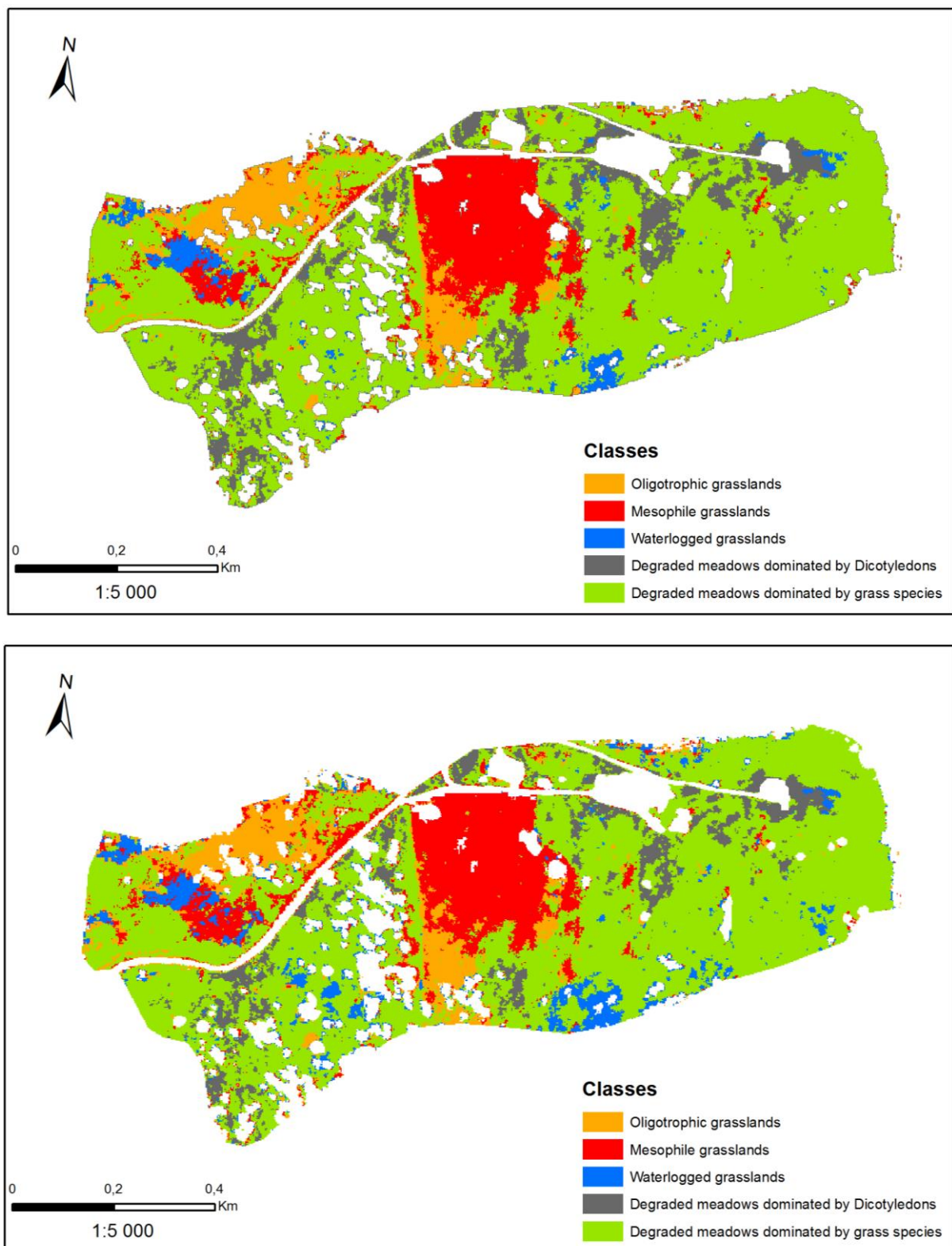


Fig. 35: Classification maps for Přední Rennerovky for randomly sampled 10% of total amount of pixels (upper) and clustered 5% of total amount of pixels (lower).

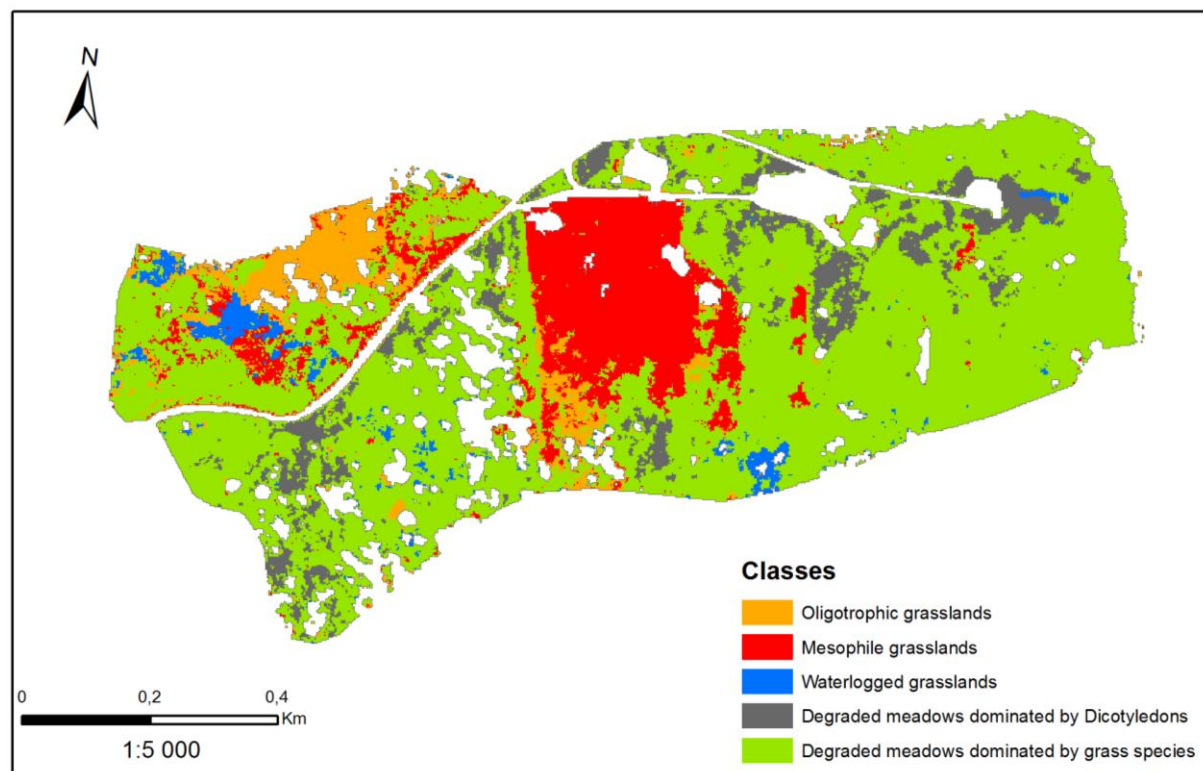
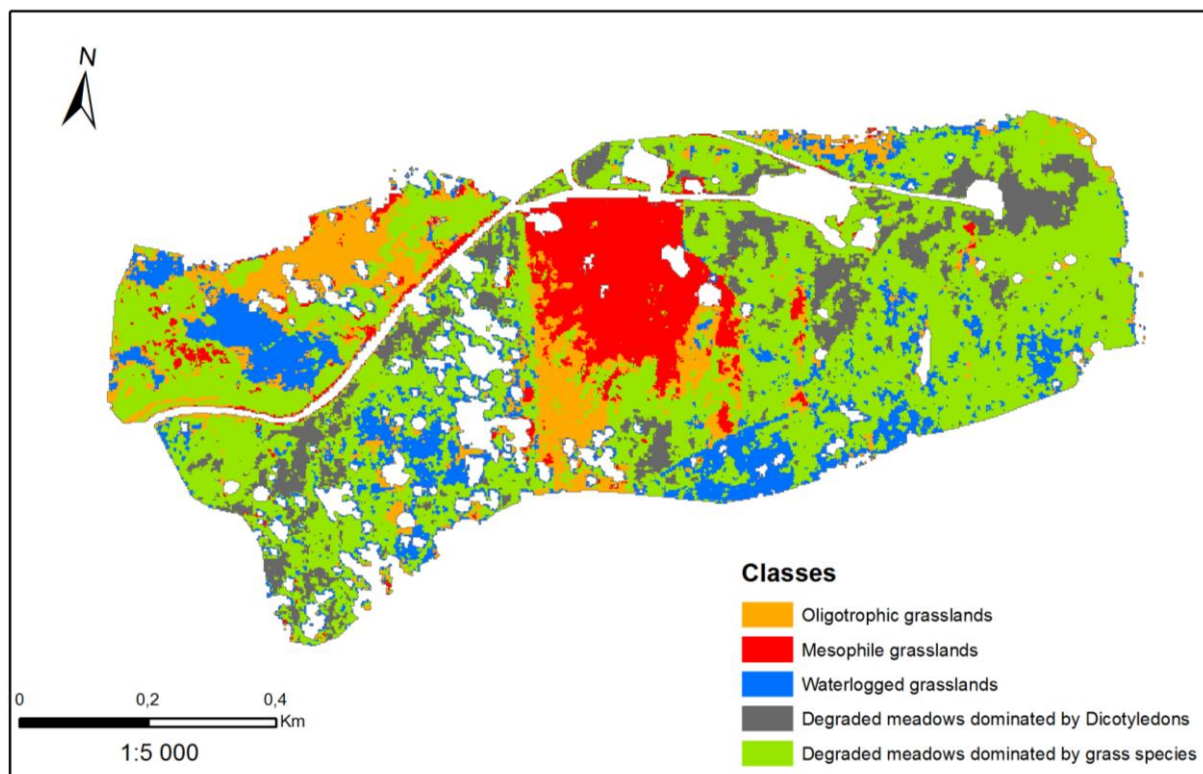


Fig. 36: Classification maps for Přebudov for “real life scenario” 50 pixels (upper) and 10% pixels at boundaries (lower).

7. Discussion

7.1 The utilized datasets and their limitations

As it was mentioned before, some errors in the aerial dataset were observed during the pre-processing of this dataset. The discrepancies in the spectral signatures, majorly in the infrared part of spectrum of the same pixel between different flight lines were considered as the most serious errors. These errors could have been caused for example by incorrect calibration of sensor between sensing of different flight lines, by wrong atmospheric corrections of particular flight lines or just by difference of one hour in sensing of the lines. A difference of one hour usually does not mean much, when images are sensed around noon. However, as some of the images were sensed in the morning and others shortly before noon, the change in the strength of sunlight could have been significant enough to cause the discrepancies in the spectral signatures. The lower values in the infrared spectrum also suggest that wrong scaling factor could have been used when DNs in the infrared part of spectrum for one flight line were converted to reflectance values. Normally scaling factor of 40 should be used for VNIR part of AISA spectrum and scaling factor of 80 should be used for SWIR part of AISA spectrum (Imaging spectroscopy, 2013), but it is possible that scaling factor of 40 has been used for the whole spectrum of this particular flight line, which caused the differences in spectra between lines. Although it is difficult to learn a real cause for the discrepancies because only limited information about the pre-processing of the aerial data was submitted by the outsourcing company.

The differences in the spectra were also the reason, why all the utilized flight lines were processed separately without mosaicking. While it was probably a correct step (because for example both parts of enclave Lahrový Boudy were classified with quite a similar accuracy and this might not have happened if the lines had been connected together), it also posed some problems mainly during the assessment of accuracy of the classifications, when lines had to be assessed also separately. Furthermore non-mosaicking probably underlined the geometric distortions on the edges of lines and in some cases caused several metres up to 20 metres wide “belts” of wrongly classified pixels along the edges of lines (appendices XI, XIII, XV and XVII). These problems are usually diminished during mosaicking because distorted pixels are balanced by undistorted pixels from middle parts of the mosaicked lines (Wang et al., 2008) and also there are no edges of lines to be misclassified. However, as mosaicking

was not performed on the utilized dataset, the distorted and wrongly classified pixels on the edges probably added to lower accuracy of the classification of some enclaves.

What's more, another shortcoming of the utilized aerial dataset was the early day time of acquisition of some images. There are huge shadows present in the images of some enclaves (especially Husí Boudy and Lahrovy Boudy), which make many pixels impossible to use for the analysis, as there is no information present in these shadowed pixels. In the case of Husí Boudy shadows made it even impossible to classify class 7 (stands dominated by *Vaccinium* species) because the vegetation of this class occurs close to a forest, right on the spots which were shadowed during the time of acquisition. Thus it was not possible to collect any pixels for training or validating.

The size of the hyperspectral dataset could be perceived as a down-side, nonetheless it was manageable with some work planning and using of spatial and spectral subsets (the original size of the images 600GB per flight line was in the end reduced to hundreds of MB).

Even if the quality of images is important, training and validating polygons have quite often higher impact on the accuracy of classification (Jones and Vaughan, 2010), as it probably happened in the case of the classifications performed in this master thesis. The field survey was done with no use of any GPS device and also the characteristics of vegetation (e.g. measure of waterlogging or percentual amount of *Dicotyledons*) were not measured but only mapped by "a human eye". These subjective ways of mapping can cause biases in the data, spatial or factual. The spatial biases have been observed during the pre-processing of the field mapping data and were corrected within the limits (e.g. misplaced polygons of roads or cottages), however different types of the meadows could not be corrected only by using orthophoto, as they are not distinguishable in it. The new field survey had to be conducted where enclave of Přední Rennerovky and partially also enclave of Friesovy Boudy were re-checked. Unfortunately no GPS device was utilized also during the new field survey, because of the lack of the material and personal resources. Nevertheless the re-mapping yielded the new set of more spatially accurate polygons, which could be used for the experiments with parameters of SVM classifier and the amount and sampling design of training pixels. Furthermore the re-checked enclaves were classified with higher accuracy than the other ones.

Last but not least is the fact that the field mapping was conducted one year later after the sensing of images. Thus the placement of some faster-growing vegetation communities (e.g. *Rumex alpinus*) could have been slightly changed. The classifiers were then fed with

wrong information about vegetation and this introduced artificial errors in the accuracy of classifications, not caused by the classification algorithms themselves. However such cases should not be common as the vegetation of meadows in the Krkonoše mountains is usually quite stable and different communities do not change their placements (consultations with Stanislav Březina).

7.2 The tested parameters of the SVM classifier

The testing of the SVM parameters was one of the main purposes of this master thesis. It was learnt that the polynomial kernel was the best suited for the classification of the meadow vegetation in the Krkonoše mountains. These findings are opposite to what was anticipated, as the RBF kernel was stated as the best in the majority of the studied literature. However Dixon and Candade (2008) and Camps-Valls et al. (2004) also mention polynomial kernels as the best performing kernels regarding the vegetation they classified. Both groups of authors classified rural areas, one in Florida (Dixon and Candade, 2008) and the other in Barrax, Spain (Camps-Valls et al., 2004). The meadows in The Krkonoše mountains are usually held under some management (e.g. grazing, mowing) (Štursa, 2013) like crops in the fields and therefore the structure of the vegetation there could remind the one of crops. Furthermore categories classified by Dixon and Candade (2008) such as wetlands or pastures or mountainous area classified by Camps-Valls et al. (2004) also resemble the landscape in the Krkonoše mountains. Moreover Dixon and Candade (2008) utilized Landsat 5 TM data with 6 input bands and Camps-Valls et al. (2004) reported polynomial kernel as the best when 6 bands of HyMap imagery were utilized. In this master thesis 7 best bands of AISA imagery were utilized, which is really similar amount to the one which Dixon and Candade (2008) and Camps and Valls (2004) utilized in their experiments, so perhaps also from this reason the same type of a kernel could give the best results. As for the polynomial degree of the polynomial kernel, the highest possible degree (6) gave the best results. This finding is fully in agreement with the studied literature (e.g. Camps-Valls et al. 2004; Huang et al. 2002), where authors stated, that accuracy of classification increases with the increasing polynomial degree of the polynomial kernel.

Regarding penalty parameter C, the highest accuracies were measured for the maximum possible value (100) of this parameter. Such result is the same as the results of studies of Petropoulos et al. (2012), Dixon and Candade (2008) or Pal and Mather (2005).

Where also maximum values allowed by utilized SW were used. According to Dixon and Candade (2008), greater value of penalty parameter yields more accurate classification results because the SVM algorithm is enabled to search in a wider space and thus classification errors are avoided by allowing more slack variables, further from a separating hyperplane. Moreover higher values of the penalty parameter are also recommended when noisy data is utilized (ČVUT, wiki course). The aerial dataset used in this master thesis can be considered as noisy, because of the high within class variability of the mountainous meadows and also because of the inexact field mapping (chapter 4.2.2) and therefore it is logical that the highest accuracies were obtained for the highest values of penalty parameter.

It was also learnt that the changes in values of gamma parameter did not have any effect on the accuracy of classification. Quite a similar behaviour was observed only in one of all the studied cases, by Huang et al. (2000) where increasing values of gamma parameter beyond value 7.5 did not significantly change the results of classification, however Huang et al. (2000) did not mention any possible reasons behind such behaviour. In my opinion, the changes in values of gamma parameter did not influence the accuracy of classifications because the potential of hyperspectral data was not fully utilized (as less detailed legend had to be used in the end that what was planned in the beginning because of the coarse field mapping) and using penalty parameter together with any non-linear transformation into higher dimensional space was enough to classify the given datasets according to a given legend.

7.3 Sampling design and the amount of training pixels

Another main task of this master thesis was to test an influence of different types of training datasets on the accuracy of classification. The initial hypothesis to test was the finding of Foody and Mathur (2004, 2006) that the SVM algorithm is able to classify datasets with the same or higher accuracy when border training pixels are used instead of the conventionally chosen ones (e.g. randomly sampled pixels). Results confirming this hypothesis were indeed observed – the accuracy yielded by using pixels at boundaries as training dataset was comparable with the accuracy gained when using randomly sampled training pixels. Such results were probably obtained because more of the support vectors, which are used by the SVM algorithm to construct a separating hyperplane, are actually situated at the boundaries of polygons than in their centres (Foody and Mathur 2004, 2006). The fact, that randomly sampled training pixels still gave better accuracies by 1 - 2%, is

probably caused by high within class variability of mountainous meadows, because, thanks to such variability, support vectors are probably contained also deeper in the polygons of these meadows, further from their boundaries. Furthermore inexact field mapping caused that small polygons of one land cover type are situated within the polygons marked as different land cover type, which artificially increases within class variability (regarding support vectors). The support vectors can be then really found deep inside of the polygons. This finding could make field mapping of hardly accessible biotopes (like peat bogs, marshes or very closed forest canopies) easier, because according to the results of the analysis, these biotopes could be mapped just on their boundaries and it would not be necessary to wade or getting through such biotopes.

Besides the two already mentioned sampling designs, other ones were tested too (the clustered sampling design and “real life scenario” sampling design). The worst accuracies were measured for the “real life scenario” sampling design endorsing the statements of Jones and Vaughan (2010) and Huang et al. (2002) that training pixels placed too close to each other bear the same correlated information and thus cannot yield sufficient accuracies. Therefore it is not recommended to use “real life scenario” sampling design, even if it is the least demanding sampling design when material and human resources are considered. The results gained by using clustered sampling design were more satisfactory than those from “real life scenario” sampling design and equivalent with results from randomly sampled and boundary training pixels. The reason for this is probably, that there were enough of clusters and those were not too large, so the spectral information contained in these clusters was various enough and support vectors could have been found properly. This opinion has been confirmed by a trend found in the results, where accuracy increases with increasing amount of clusters (because of the increasing amount of non-correlated samples) and oppositely it decreases with increasing size of clusters. The number of clusters was from practical reasons assigned as 1% of all the pixels in a class, however this rule could not be applied when 50, 100 and 200 training samples were used, because there would be too many centres of clusters. Therefore the amount of centres was set manually to 3, 6 and 12 for 50, 100 and 200 training pixels respectively. And the amount of pixels was then chosen from 24-N neighbourhood. While the same neighbourhood was utilized also for 10% of total amount of pixels in a class, the pixels in the case of 2.5% and 5% of total amount of pixels in the class were chosen from 8-N neighbour. The increase in the size of the clusters instead of increasing of their amount

could cause the drop in the accuracy of classification from 82.41% (for 5% of total amount of pixels in a class) to 82.11% (for 10% of total amount of pixels in a class) (table 8). The trend of increasing accuracy with the increasing amount centres of clusters can be also observed in the table 8 (58.19% for 3 centres through 71.11% for 12 centres up to 82.42% for 1% of centres, which was more than 12 in the majority of cases).

Regarding the amount of training samples, the only conclusion which could be done was the one that with the growing amount of training pixels also the accuracy of classification rises. Probably since a larger training dataset has a greater chance of including support vectors that define the actual decision boundaries and hence should give higher accuracies (Huang et al. 2002). Although as the SVM classifier uses only a portion of the training dataset (support vectors) this does not have to be valid in all the cases (Foody and Mathur, 2004). Moreover equal sample rate sampling scenarios gave better overall accuracies than equal sample size sampling scenarios. Nevertheless, when confusion matrices were examined in the depth it has been shown, that when ESR scenarios underestimated less abundant classes and pixels of these classes were usually assigned to the class with the most pixels for both SVM and NN algorithms. This effect was even stronger when differences between amounts of training pixels were several orders (e.g. tens of thousands versus few tens of pixels). This behaviour was observed also by Huang et al. (2002) not only for the SVM algorithm but also for the NN algorithm. Andrštová (2014) observed such behaviour only for the NN algorithm. What's more she states that this problem is of minor importance when aerial data with finer spectral and spatial resolution like AISA are utilized. Based on the results of this master thesis it could be said, that the importance of this problem is more significant with the growing differences between quantities of training pixels of particular classes. Consequently it would be benefiting to use a hybrid between ESR and ESS sampling scenarios in the future. For example, classes could be divided into several groups according to the total area that is covered by them. Then certain amount of pixels would be sampled from each class, similar for all the classes in one group. In this way, the variability of the classes regarding their size would be encompassed while maintaining the right differences (not too small and not too huge) between amounts of pixels of different classes.

The only case when ESS sampling scenario was more successful than the ESR one was the combination or “real life scenario sampling design” and ESS sampled pixels. This was also the only situation when “real life scenario” yielded better results than the other

sampling scenarios. Most likely because the less correlated pixels the better and the redundant information is multiplied the least in this case.

Even if an ideal training dataset could not be determined with certainty, the best combination of amount and sampling design of pixels was 10% of randomly sampled pixels, therefore this combination was thought of as the ideal training dataset and it was used for the classification of the remaining enclaves and also for the comparison of SVM and ANN algorithms.

7.4 Results of classification

The best accuracy of classification was measured at re-mapped enclaves, which is understandable because many mistakes made during the original inexact field survey were corrected. Locality Friesovy Boudy had the highest overall accuracy of all the locations, however 2 out of the 4 total classes were not classified at all. These classes (class 4 and class 7) were less abundant than other ones (by few orders) and that was probably the reason why they did not get classified. The classification of enclave Přední Rennerovky (the model enclave) reached OA of 83.63% and with majority of the classes classified correctly in at least 50% of cases this enclave can be considered as the most accurately classified. Nonetheless class 4 (waterlogged grasslands) had the lowest accuracy of classification also at this enclave (45.05%). Another reason for worse classification accuracy on non-checked enclaves might be the wrong categorization. Unfortunately categorization of vegetation was not done during the field mapping but few months later and based on only few notes, thus it is highly possible that some polygons of less distinguishable meadows vegetation were mixed between each other into wrong classes. Such classes could be for example class 1 and class 2 where only the ratio of *Nardus stricta*, which was not usually mentioned in the notes, to other vegetation in the polygon makes the difference between these classes (polygons with dominant *N.stricta* are assigned to class 1 and others to class 2).

Also it could be said that in general the more spacious enclave the worse accuracy of classification. The reason for this behaviour is that the each enclave was mapped on one sheet of A4 paper, which means that the more spacious enclaves were mapped with a coarser scale. However as the field survey was conducted by a hired botanist (chapter 4.2) and we obtained only results of the field mapping, it could not be prevented. The only exception in this is the part of Zadní Rennerovky on flight line 10_20, probably because land cover in this upper part

of Zadní Rennerovky is less diverse than in the lower part (flight line 10_58) and rather lower amount of bigger polygons than high amount of smaller ones occur here. Such bigger polygons are then easier to map even with a coarser scale.

As anticipated, the classes which were more abundant on a certain location were mapped with higher accuracy in comparison with other classes on this location. The degraded meadows with grass species (class 6) are the most common land cover type on the majority of mapped locations and most likely because of this they were mapped with the highest accuracy (usually over or around 90%) on these locations (because of the behaviour of SVM and NN classifiers mentioned in the previous chapter). The high accuracy of the classification of such abundant class (usually over 100 thousands classified pixels and almost never under 50 thousands classified pixels) understandably adds significant percentage also to the overall accuracy of classification, therefore it could be anticipated that the overall accuracy of classification was increased by the presence of this abundant class. There were 3 enclaves (Husí Boudy, Lahrový Boudy on both lines) with classes other than class 6 classified with the highest accuracy. These enclaves are smaller and therefore easier to manage (e.g. cut or graze) therefore degraded grasslands are not the most abundant class and other classes take their spot. Oligotrophic and mezophile grasslands are the most common on the enclave of Husí Boudy and therefore also the best classified category (oligotrophic grasslands in the case of SVM classifier with accuracy 73% and mesophile grasslands in the case of NN classifier with accuracy 82%). The most abundant class for Lahrový Boudy on both flight lines are mesophile grasslands which were also the best classified category by SVM classifier also on both flight lines (accuracy 79% and 86%). The NN classifier mapped the best class 1 (*Nardus stricta* grasslands) on flight line 10_45 (accuracy 79%) and class 5 (degraded grasslands by *Dicotyledons*) on flight line 08_45 (accuracy 80%), these classes are also quite abundant on this enclave. The difference between mapped classes by NN classifier between the flight lines could be caused by the difference in spectral curves of these lines mentioned in chapter 4.1.3 and shown in fig. 24. SVM classifier seems to be more robust in this case and not to be much influenced by different ratios of shadowed parts of vegetation to non-shadowed ones.

The worst classified classes were classes 4 (waterlogged grasslands; accuracy 0 to 37%), 7 (stands with *Vaccinium* species; accuracy 0 to 38%), 8 (springs; accuracy 0 to 19%) and 9 (tall-fern vegetation; accuracy 0 to 0.53%). The class 9 was properly classified at neither of two enclaves (Klínové Boudy and Zadní Rennerovky) where it occurs. It could be

either because of its low abundance or because of the coarse scale under which these enclaves were mapped. Other wrongly classified classes are also of low abundance in general. Furthermore they also get mixed with other classes. Class 4 was often mixed with classes 5 and 6, because these vegetation communities can also grow on locally waterlogged areas (consultations with Stanislav Březina). Class 8 got often mixed with class 6 from the same reason mentioned in the previous sentence. In my opinion it would be possible to prevent some misclassifications of waterlogged vegetation or springs if measure of waterlogging was quantified by exact methods instead of by “a human eye” as this could also introduce some bias into classification. As *Vaccinium* species also belong to *Dicotyledons*, class 7 got often mixed with other classes where *Dicotyledons* occur abundantly (e.g. class 5) or with classes creating mosaic with this class (e.g. class 6 or class 1). Also class 3 and class 5 got often confused between each other, because there is high ratio of *Dicotyledons* in both of these classes. Moreover *Dicotyledons* of mesophile grasslands can sometimes create clusters, where less grass species is contained and these pixels might then be assigned to class 5 instead of class 3. The same situation can happen also in the vegetation communities of class 5, where clustering of *Dicotyledons* can oppositely cause creation of spots with local higher abundance of grass and these pixels are then assigned to class 3 instead of class 5. Classes 6, 1 and 2 were occasionally confused amongst themselves too. These confusions happened most likely because of the dominance of grass species in all three classes. What’s more these classes often verge between each other and there are no clear boundaries between them (consultations with Stanislav Březina), which also makes it difficult for a classifier to assign pixels properly into these classes.

In general, SVM slightly outperformed NN classifier regarding classification accuracy (average accuracy 70% achieved by SVM classifier versus average accuracy 68% achieved by NN classifier) however with up to three times shorter computational times. The overall accuracies yielded by NN classifier were normally 2 to 4.5% lower than the ones yielded by SVM classifier. Kappa coefficients were lower by 0.02 to 0.06. Usually, there were also more classes left unclassified by NN classifier than by SVM (especially on enclaves Zadní Rennerovky and Klínové Boudy). The only exception was enclave of Friesovy Boudy where both classifiers performed comparably. These results are in agreement with the studied literature (table 4) and show that SVM classifier might be more suitable for the classification of such diverse land cover type as mountainous meadows. Moreover accuracies of

classification achieved by NN classifier are comparable to those achieved by SVM classifier when same sampling design but less pixels are utilized (accuracy 80.37% for NN classifier and randomly sampled 10% of total amount of pixels is quite similar to accuracy 81.80% for SVM classifier when randomly sampled 2.5% of total amount of pixel was used). It could be then stated that SVM needs less training pixels than NN classifier to achieve comparable accuracy of classification, probably because of the way in which SVM finds a separating hyperplane with use of the support vectors and the possibility of transforming feature space into higher dimensional space where classes can be efficiently separated.

7.5 A comparison of the results of classification with results from other studies

The meadows of the Krkonoše mountains have been previously classified in several master theses mentioned in chapter 2.4. The master thesis of Dorič (2013) was considered as the main reference to this master thesis, because of quite similar legend and SVM classifier utilized. However it was anticipated that the accuracy of classification achieved by Dorič would be lower because of the use of multispectral data. This hypothesis was confirmed (the highest overall accuracy around 85% versus 59% achieved by Dorič) and thus it was confirmed that the hyperspectral data enable us to better distinguish complex mountain meadow vegetation communities, like those in the Krkonoše mountains. The highest overall accuracies gained by the methods utilized in this master thesis was also higher than the accuracies achieved by Jelének (2013) and higher than some of the accuracies achieved by Pomahačová (2012). This result shows that a combination of SVM classifier and hyperspectral data might be more suitable than combinations of methods and datasets utilized in theses of Jelének (2013) and Pomahačová (2012). However it is necessary to keep in mind that the items classified in their theses were a bit different than those classified in this master thesis and thus results are not fully comparable. Biotop quite similar to mountain meadows (subalpine tundra) was classified also by Andrštová (2014) and it was classified with the higher accuracies than those achieved in this master thesis (highest overall accuracies achieved by Andrštová were more than 99% by classifiers SAM and LSU). Nonetheless the boundaries between vegetation types of tundra are usually defined much sharply than those between different vegetation of mountain meadows (Andrštová 2014), and land cover types are thus easier to classify. This fact can be also seen when we compare master theses of

Jelének and Andrštová both of who utilized hyperspectral data from sensor APEX and SAM classification method, however on slightly different land cover type and with huge differences in results. While SAM was denoted by Andrštová (2014) as appropriate for classification of subalpine tundra, it was considered as unsuitable by Jelének (2013) for classification of mountain meadows. Andrštová (2014) also mentioned that classification accuracies carried out by NN classifier were higher than those carried out by SVM classifier, which is the opposite statement to the one posed in this master thesis. Such diverse conclusions only endorse the fact that even slight changes in a structure or other characteristics of otherwise similar biotopes can have significant impact on classification results of these biotopes.

As for the comparison of SVM classification accuracies with those achieved by SVM worldwide (table 4), they were generally better than those gained by utilizing of multispectral data and average with regard to those achieved by use of hyperspectral data. Alike trend can be observed also when we compare the results yielded by NN classifier (table 4). Even though the results achieved in this master thesis were mediocre in comparison with those achieved worldwide, the classifications can still be considered as successful, especially with respect to many shortcomings of aerial and field mapping datasets utilized. Especially the quality of the field mapping dataset is considered as the major drawback which reduced the accuracy of classifications, because as it was proved in the study of Marcinkowska et al. (2014), it was possible to classify even individual species alliances of meadows in the Krkonoše mountains with higher producer and user accuracies (in majority over 80% and in many cases even over 90%) when hyperspectral data and SVM algorithm were used and training and validating samples were properly measured by a GPS device.

8. Conclusion

Mountainous grasslands in the Krkonoše mountains have been classified in this master thesis with use of aerial hyperspectral data from AISA DUAL sensor and SVM classification algorithm. One of the two main goals of this master thesis was to experimentally test which setting of parameters of SVM classifier is the best for the given task. Therefore series of trials were performed on the model location of Přední Rennerovky. In these trials several kernels (polynomial, RBF and linear) were tested with various combinations of gamma and penalty parameters and in the case of polynomial kernel also with different values of polynomial degree. Polynomial kernel of 6th polynomial degree and the maximum value of penalty parameter (value of 100) gave the best results of classifications. Consequently this combination of parameters of the SVM classifier was utilized to achieve 2nd main goal of this master thesis – to determine ideal sampling design and amount of training pixels for the classification of mountain meadows and also to test the initial hypothesis. The initial hypothesis to test was the finding of Foody and Mathur (2004, 2006) that the SVM algorithm is able to classify datasets with the same or higher accuracy when border training pixels are used instead of the conventionally chosen ones (e.g. randomly sampled pixels). Again, series of classification trials were conducted with in total 24 combinations of sampling design and amount of pixels, of which random sampling design yielded the highest overall accuracies. Nevertheless also pixels sampled from boundaries of polygons gave satisfactory results (only up to 2% lower than randomly sampled pixels), therefore it could be stated that the hypothesis of Foody and Mathur was confirmed and thus this sampling design could be utilized in the cases when random sampling is difficult to perform (e.g. on peat bogs, springs or other heavily waterlogged areas) to ease the collecting of training samples. Unfortunately the exact minimum amount of training pixels could not be determined. However, even if the best results were gained for the highest amount of pixels utilized (10% of total amount of pixels), accuracies over 80% were measured also when only 2.5% of total amount of pixels per class was used. Furthermore it was learnt that equal sample rate sampling strategy, where proportions of particular classes occurring on enclaves are represented better, yielded more generally satisfactory results than equal sample size strategy, where same amount of pixels is collected for each class regardless its abundance on a location. Nonetheless equal simple rate strategy is not the best possible sampling strategy, because huge differences in amount of pixels between certain classes are sometimes created by this strategy, especially when there

are both over-abundant and not numerous classes on a location. Therefore a hybrid between equal simple rate and equal simple size was proposed as the best sampling strategy regarding amount of training pixels.

An additional goal of this master thesis was to compare results of classification achieved by SVM classifier and the ideal training dataset with results of classification gained by NN classifier and the same training dataset. SVM classifier generally outperformed NN classifier regarding accuracy of classification and also computational time. Moreover accuracies achieved by NN classifier were comparable to those achieved by SVM classifier when same sampling design (random sampling design) but less training pixels were utilized (2.5% of total amount of pixels). Therefore it could be said that SVM needs less training pixels than NN classifier for the classification of given biotope and under given conditions.

In the course of this master thesis, during the pre-processing of aerial and field mapping datasets, it was also necessary to cope with several inconveniences. At first, some enclaves were divided between two flight lines, which could not be mosaicked due to some discrepancies in spectral features of particular lines. Secondly coarse field mapping caused spatial misplacements of mapped polygons and those had to be corrected. However as the field mapping was not conducted with the use of a GPS device, buffer of -2 metres was applied to all polygons enclaves to deal with possible mapping inaccuracies. What's more, while the factual information of polygons was contained in MS Excel tables, the spatial information about these polygons was contained in shapefiles. Both of them had to be joined and also the mask of non-classified areas had to be applied on the shapefiles. All the mentioned inconveniences were dealt with by a single Python script, which eased otherwise difficult and time-consuming work.

The best accuracies were achieved on locations of Přední Rennerovky and Friesovy Boudy, 83.63% and 87.63% for SVM classifier respectively, and 80.37% and 87.35% for NN classifier respectively. These results are considered as satisfactory because the meadows of the Krkonoše mountains were mapped with higher accuracies than in the most of previous cases. An improvement has been made especially in comparison with Dorič's master thesis, where the highest achieved accuracies were under the 60%, the increase in accuracy was then more than 20%. The use of hyperspectral data was accounted for this increase in accuracy of classifications and thus it was confirmed that the data with higher spectral (and spatial) can significantly improve the results of classification of mountainous meadows.

All of the goals of this master thesis were fulfilled and the results of the analyses together with Python scripts will be handed over to the administration of the Krkonoše Mountains National Park for further use in mapping and managing the sensitive biotope of mountainous meadows.

Further continuation of this work could focus on classifying the same dataset with the use of a hybrid sampling rate as was proposed in the chapter 7.3 or with the “boundary training samples” defined on boundaries that are not given geographically (as the division line of different class polygons) but based for example on the change of factors which influence spatial distribution of single meadow communities (e.g. type of the soil, slope or orientation of the slope).

9. References

AISA, 2013 – Hyperspectral scanning, vegetation maps and inventory of forest stands; KRNAP technical report for year 2013;

AISA EAGLE – datasheet, online at:

http://www.specim.fi/files/pdf/aisa/datasheets/AisaEAGLE_datasheet_ver1-2013%281%29.pdf (last viewed 22th July 2015)

AISA DUAL – datasheet, online at: http://www.specim.fi/files/pdf/aisa/datasheets/Dual_datasheet_ver1-2012.pdf (last viewed 22th July 2015)

AISA HAWK – datasheet, online at:

<http://www.channelsystems.ca/documents/AISAHawkver1-07.pdf> (last viewed 22th July 2015)

Andrštová, M., 2014; Využití hyperspektrálních dat ke klasifikaci vegetace alpského bezlesí v Krkonoších; master thesis (Department of applied geoinformatics and cartography; Faculty of science; Charles university of Prague);

APEX-ESA; online at: <http://www.apex-esa.org/> (last viewed 22th July 2015)

Belousov A.I., Verzakov S.A., von Frese J., 2002, A flexible classification approach with optimal generalisation performance: support vector machines; chemometrics and Intelligent laboratory systems, 64. 15 – 25;

Benediktsson J.A, Swain P.H., Ersoy O.K; 1990, Neural network approaches versus statistical methods in classification of multisource remote sensing data; IEEE Transactions on geoscience and remote sensing, 28/4, 540 – 551;

Borengasser M., Hungate W.S., Watkins R.; Hyperspectral remote sensing: principles and applications; Taylor&Francis, 2008,

Březina, Stanislav; specialist of Environmental protection department of the Krkonoše Mts. National Park Administration

Campbell, J.B., Wynne, R.H., Introduction to remote sensing, 5th edition, The Guildford press, 2011, ISBN 978-1-60918-176-5;

Camps-Valls, G., Gómez-Chova L., Calpe-Maravilla J., 2004, Robust support vector method for hyperspectral data classification and knowledge discovery, IEEE Transactions on geoscience and remote sensing, 20, 1 – 13;

Chan J.C-W., Beckers P., Spanhove T., Borre J.V.; 2012, An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery; International Journal of Applied Earth Observation and Geoinformation, 18, 13 – 22;

Chen C.F, Chen C.R., Son N.T, Chang L.Y.; 2012, Delineating rice cropping activities from MODIS data using wavelet transform and artificial neural networks in the Lower Mekong countries; Agriculture, Ecosystems and Environment, 162, 127 – 137;

Chytrý M., Kučera T., Kočí M. (eds.), 2001, Katalog biotopů České republiky. Agentura ochrany přírody a krajiny ČR, Praha

ČÚZK, ortophoto; online at:

<http://geoportals.cuzk.cz/%28S%28ep11a2bxap5w5dt2qw0g1ycb%29%29/Default.aspx?mode=TextMeta&metaDataXSL=full&side=wms.verejne&metadataID=CZ-CUZK-WMS-ORTOFOTO-P> (last viewed 22th July 2015)

ČVUT, wiki course; online at: https://cw.fel.cvut.cz/wiki/courses/ae4b33rpz/labs/07_svm/start (last viewed 22th July 2015)

Debeir O., Van den Steen I., Latinne P., Van Ham P., Wolff E., 2002; Textural and contextual land-cover classification using single and multiple classifier systems, Photogrammetric engineering and remote sensing, 68, 597 – 605;

Demir, B., and Ertürk, S., 2009; Clustering-Based Extraction of Border Training Patterns for Accurate SVM Classification of Hyperspectral Images, IEEE Geoscience and Remote Sensing Letters, 6/4, 840-844

Dixon B., Candade N., 2008; Multispectral landuse classification using neural networks and support vector machines: one or the other, or both?; International Journal of Remote Sensing, 29/4, 1185-1206

Dobrowski S.Z., Safford, H.D., Cheng, Y.B., Ustin, S.L.; 2008, Mapping mountain vegetation using species distribution modelling, image-based texture analysis and object-based classification; Applied vegetation science, 11(4), 499 – 508;

Dorič, R., 2013; Možnosti objektovo orientovanej klasifikácie při monitoringu lúčnej vegetácie a manažmentových zásahov v Krkonošskom národnom parku; master thesis (Department of applied geoinformatics and cartography; Faculty of science; Charles university of Prague);

ENVI, MNF; online at: <http://www.exelisvis.com/docs/MinimumNoiseFractionTransform.html> (last viewed 22th July 2015)

ENVI, NN; online at: <http://www.exelisvis.com/docs/NeuralNet.html> (last viewed 22th July 2015)

ENVI, PCA; online at: <http://www.exelisvis.com/docs/PrincipalComponentAnalysis.html> (last viewed 22th July 2015)

ENVI, SVM; online at: <http://www.exelisvis.com/docs/SupportVectorMachine.html> (last viewed 22th July 2015)

ENVI, SVM background; online at: <http://www.exelisvis.com/docs/BackgroundSVM.html> (last viewed 22th July 2015)

Feilhauer H., Asner G.P., Martin, R.E., Schmidtlein S., 2010, Brightness normalized partial least square regression for hyperspectral data, Journal of Quantitative Spectroscopy and Radiative Transfer, 111, 1947 – 1957;

Foody G.M., Mathur, A., 2004; Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification, Remote sensing of environment, 93, 107 – 117;

Foody G.M., Mathur A., 2006; The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM, *Remote Sensing Environment*, 103/2, 179-189;

Huang, C., Davis, L.S., Townshend, J.R.G, 2002, An assessment of support vector machines for land cover classification; *International journal of remote sensing*, 23/4; 725 – 749;

Imaging spectroscopy, 2013; lectures from subject Imaging Spectroscopy

Jakešová L., 2014; Možnosti objektově-orientované klasifikace pro určování vybraných biotopů nad horní hranicí lesa v Krkonošském národním parku; bachelor thesis (Department of applied geoinformatics and cartography; Faculty of science; Charles university of Prague);

Jakomulska, A., Zagajewski B., Sobczak, M., 2003; Field remote sensing techniques for mountains vegetation investigation; 3rd EARSEL Workshop on Imaging Spectroscopy; 13th-16th May 2003;

Jelének J., 2013; Vytvoření algoritmu klasifikace vybraných invazivních druhů a lučních společenstev v Krkonoších s využitím hyperspektrálních dat; master thesis (Department of applied geoinformatics and cartography; Faculty of science; Charles university of Prague);

Jensen, J.R.; Remote sensing of the environment: an earth resource perspective, 2nd edition, Upper Saddle River, 2007, ISBN 978-013-1889-507;

Jones H.G. Vaughan R.A.; Remote sensing of vegetation: Principles, techniques and applications; Oxford: Oxford university press, 2010, ISBN 978-0-19-920779-4;

Karvánek M., 2014; Hodnocení změn krajiny v Krkonošském národním parku s využitím dálkového průzkumu země a krajinných metrik; bachelor thesis (Department of applied geoinformatics and cartography; Faculty of science; Charles university of Prague);

Krahulec, F., Blažková, D., Pecháčková, S., Fabšičová, M., Balátová-Tuláčková, E., Štursa, J.; 1997; Louky Krkonoš: Rostlinná společenstva a jejich dynamika. *Opera Corcontica*, 33, 3- 250;

KRNAP, climate; online at: <http://www.krnep.cz/podnebi/> (last viewed 22th July 2015)

KRNAP, flora; online at: <http://www.krnep.cz/flora-a-vegetace/> (last viewed 22th July 2015)

KRNAP, precipitation; online at: <http://www.krnep.cz/srazky/> (last viewed 22th July 2015)

KRNAP, snow; online at: <http://www.krnep.cz/snih-a-laviny/> (last viewed 22th July 2015)

KRNAP, sunlight; online at: <http://www.krnep.cz/oblacnost-a-slunecni-svit/> (last viewed 22th July 2015)

KRNAP, temperature; online at: <http://www.krnep.cz/teplota/> (last viewed 22th July 2015)

KRNAP, wind; online at: <http://www.krnep.cz/vetrne-proudeni/> (last viewed 22th July 2015)

Lucas, R., Rowlands, A., Brown, A., Keyworth, S., Bunting, P.; 2007; Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping; *ISPRS Journal of photogrammetry & Remote Sensing*, 62; 165 – 185;

Marcinokowska, A., Zagajewski, B., Ochtyra, A., Jarocińska, A., Raczko, E., Kupková, L., Štych, P., Meuleman, K., 2014, Mapping vegetation communities of the Karkonosze National Park using APEX hyperspectral data and Support Vector Machines; *Miscellanea Geographica – Regional Studies on Development*; 18/2, 23-29;

Magiera A., Feilhauer H., Otte, A., Wadhardt, R., Simmering, D.; 2013, Relating canopy reflectance to the vegetation composition of mountainous grasslands in the Greater Caucasus; *Agriculture, Ecosystems and environment* 177, 101 – 112;

Melgani F., Bruzzone, L., 2004; Classification of Hyperspectral Remote sensing images with support vector machines; *IEEE transactions on geoscience and remote sensing*, 42/8; 1778 – 1790;

Meuleman, K., 2012; Presentation on APEX – Airborne Prism Experiment

Mountrakis G., Im J., Ogole C., 2011; Support vector machines in remote sensing: A review; *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 247 – 259;

Norton P., 2002; HgCdTe infrared detectors; *Opto-electronics review*, 10(3), 159–174;

Petropoulos, G:P., Arvanitis K., Sigrimis, N.; 2012; Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping; *Expert systems with Applications*, 39, 3800 – 3809;

Pal M. and Mather P.M., 2005; Support vector machines for classification in remote sensing, *International Journal of Remote Sensing*, 26/5, 1007 – 1011;

Pal M. and Mather P.M., 2006; Some issues in the classification of DAIS hyperspectral data, *International Journal of Remote Sensing*, 27, 2895-2916

Pomahačová M., 2012; Možnosti využití DPZ pro monitoring luční vegetace a managementových zásahů v Krkonoších; master thesis (Department of applied geoinformatics and cartography; Faculty of science; Charles university of Prague)

Reddit, machine learning; online at:

http://www.reddit.com/r/MachineLearning/comments/15zrpp/please_explain_support_vector_machines_svm_like_i ; image number 7: <http://i.imgur.com/WuxyO.png> (last viewed 22th July 2015)

Roberts, D.A., Keely L.R., Ryan L.P.; 2011; Hyperspectral vegetation indices; in Thenkabail P.S., Lyon J.G., Huete, A., *Hyperspectral remote sensing of vegetation*; CRC press 2012; 309 – 327

Schmidtlein S, Sassin J., 2004; Mapping of continuous floristic gradients in grasslands using hyperspectral imagery; *Remote sensing of environment* 92, 126 – 138;

Sha Z., Bai, Y., Xie, Y., Yu, M., Zhang, L., 2009, Using a hybrid fuzzy classifier (HFC) to map typical grassland vegetation in Xilin River Basin, Inner Mongolia, China; *International journal of remote sensing*; 29:8, 2317 – 2337;

Sulzer W., Gspurning J., Magnes M., Pink R., Muick M., Sengl P.; 2013; biotope mapping of extensive/intensive grassland supported by remote sensing and mobile GIS in eastern Styria (Austria); *Ekológia (Bratislava)*, 32(4), 335 – 344

Štursa, J., 2013 Arktoalpínská tundra Krkonoš, *Živa (Academia)* 4, 171-174; online at: <http://ziva.avcr.cz/2013-4/arktoalpínska-tundra-krkonos.html> (last viewed 22th July 2015)

Vyas, D., Krishnayya N.S.R., Manjunath K.R., Ray, S.S., Panigrahy S., 2011; Evaluation of classifiers for processing Hyperion (EO-1) data of tropical vegetation; International journal of applied earth observation and geoinformation, 13, 228 – 235;

Wang, F., Huang J., Tang, Y., Wang X., New vegetation index and its application in estimating LAI of rice; 2007; Rice Science, 14, 195 – 203;

Wang, Y., Wu Y.-d., Wang H., 2008, Free image registration and mosaicking based on TIN and improved Szelisky algorithm; Remote Sensing and Spatial information sciences, 37, 145-150

Other unquoted references used for the writing of scripts for the master thesis:

ArcGIS help library (Tool reference); online at:

http://resources.arcgis.com/en/help/main/10.2/index.html#/A_quick_tour_of_geoprocessing_tool_references/002t0000000z000000/ (last viewed 22th July 2015)

Python documentation; online at: <https://docs.python.org/2/index.html> (last viewed 22th July 2015)

Stack exchange (GIS); online at: <http://gis.stackexchange.com/questions/29735/select-features-by-attribute-if-in-python-list> (last viewed 22th July 2015)

10. Appendices

Appendix I	“Inputs to classes” Python script (key part)
Appendix II	“Classes to trai-vali” Python script (key parts)
Appendix III	The classification maps of Přední Rennerovky
Appendix IV	The error matrices for the classification of Přední Rennerovky
Appendix V	The classification maps of Husí Boudy
Appendix VI	The error matrices for the classification of Husí Boudy
Appendix VII	The classification maps of Friesovy Boudy
Appendix VIII	The error matrices for the classification of Friesovy Boudy
Appendix IX	The classification maps of Klínové Boudy
Appendix X	The error matrices for the classification of Klínové Boudy
Appendix XI	The classification maps of Lahrovy Boudy on the flight line 08_43
Appendix XII	The error matrices for the classification of Lahrovy Boudy (line 08_43)
Appendix XIII	The classification maps of Lahrovy Boudy on the flight line 10_45
Appendix XIV	The error matrices for the classification of Lahrovy Boudy (line 10_45)
Appendix XV	The classification maps of Zadní Rennerovky on the flight line 10_58
Appendix XVI	The error matrices for the classification of Zadní Rennerovky (line 10_58)
Appendix XVII	The classification maps of Zadní Rennerovky on the flight line 10_20
Appendix XVIII	The error matrices for the classification of Zadní Rennerovky (line 10_20)
Appendix XIX	Contents of the attached CD

Appendix I – key part of “Inputs to classes” Python script

(source: own Python IDLE output)

```
##assign class description to each polygon
#cycle for each raster dataset (mask):
for i in MaskList:
    #dissolve polygons in each feature class (in order to avoid possible errors in case there
are several polygons with same ID - similar type of vegetation)
    #name of output dissolved feature class
    i_split = string.split(i,".")[0]
    features_id = i_split[:5]+"_id_diss.shp"

    if not arcpy.Exists(features_id):
        #name check
        try:
            value = int(i_split[-1])
            i = i_split[:-1]
        except ValueError:
            i = i_split

        arcpy.Dissolve_management(i+".shp", features_id, "Id")

        #add field "CLASS" to feature class
        arcpy.AddField_management(features_id,"CLASS","text")

        #create search and update cursors and a list for classes occurring in the area
        dbf = string.strip(i_split[:8])+"_dbf.dbf"
        cursor = arcpy.SearchCursor(dbf, sort_fields = "CISLO A")
        cursor2 = arcpy.UpdateCursor(features_id, '"Id" > 0',sort_fields = "Id A")
        classes = []

        #copy information about class from dbf.table to field "CLASS"
        for row in cursor:
            row2 = cursor2.next()

            #get information about class and in case the class is not in masked class and in
the list of classes, append it to the list
            f_class = row.getValue("TRIDA")
            if f_class not in unclassified:
                if f_class not in classes:
                    classes.append(f_class)

            #copy information
            row2.setValue("CLASS",row.getValue("TRIDA"))
            cursor2.updateRow(row2)

        del cursor, cursor2

    ##convert masks from raster to polygon and select polygons where value of the mask is
equal to 1
    in_raster = i_split+".tif"

    ##snapping (for raster to polygon conversion and create fishnet)
    arcpy.env.snapRaster = in_raster

    out_poly = i_split+"_poly.shp"
    arcpy.RasterToPolygon_conversion(in_raster, out_poly, "NO_SIMPLIFY","VALUE")

    poly_layer = i_split+"_poly.lyr"
    arcpy.MakeFeatureLayer_management(out_poly, poly_layer)
    arcpy.SelectLayerByAttribute_management(poly_layer,"NEW_SELECTION", ' "GRIDCODE" = 1 ')

    ##create fishnet for each location
    #name of output fishnet
    fishnet = i_split+"_fish.shp"

    #extent of raster
    top = arcpy.GetRasterProperties_management(in_raster, "TOP")
    left = arcpy.GetRasterProperties_management(in_raster, "LEFT")
    bottom = arcpy.GetRasterProperties_management(in_raster, "BOTTOM")
    x = float(left.getOutput(0).replace(",","."))
    y = float(bottom.getOutput(0).replace(",","."))
    origin_coord = str(x)+' '+str(y)
    x = float(left.getOutput(0).replace(",","."))
    y = float(top.getOutput(0).replace(",","."))
```

```

y_axis_coord = str(x)+' '+str(y)

arcpy.CreateFishnet_management(fishnet, origin_coord, y_axis_coord, '1',
'1','0','0',"#","NO_LABELS",in_raster,"POLYGON")

##dissolve polygons according to field "CLASS" -> there will be as many rows as there is
classes
#name of output featureclass
features_class = i_split[:5]+"_cl_diss.shp"

if not arcpy.Exists(features_class):
    arcpy.Dissolve_management(features_id, features_class, "CLASS")

##create buffered shapefiles, each of which will correspond to one class on the location,
then select pixels from fishnet which are within these layers and
##finally intersect this selection with areas where value of the mask is equal to 1

#create search cursor
cursor = arcpy.SearchCursor(features_class)
for row in cursor:
    f_class = row.getValue("CLASS")
    if f_class in classes:

        #create buffer
        #name of output buffer
        out_buffer = "buff_"+i_split[:5]+"_"+f_class+".shp"

        if not arcpy.Exists(out_buffer):
            arcpy.Buffer_analysis(row.Shape, out_buffer, -2)

        #select such pixels from fishnet which are within buffer zone
        fish_layer = i_split+"_fish.lyr"
        arcpy.MakeFeatureLayer_management(fishnet, fish_layer)
        arcpy.SelectLayerByLocation_management(fish_layer, "within", out_buffer,"#",
"NEW_SELECTION")

        #intersect the 2 previous selections, this is final output of this script
        in_features = [poly_layer,fish_layer]
        out_intersect = i_split+"_"+f_class+".shp"
        arcpy.Intersect_analysis(in_features, os.path.join(out_shp,out_intersect),
"NO_FID", "", "INPUT")

del cursor

```

Appendix II – key parts of “Classes to trai-vali” Python script

Appendix II.A. Random sampling design, an example for 10% of total amount of all pixels in a class (source: own Python IDLE output).

```
##RANDOM POINTS, % AMOUNT
#randomly choose 10% of rows
number = count*0.10
rand_IDS_10 = sorted(random.sample(IDs,int(round(number))))
rand_IDS_10 = str(tuple(rand_IDS_10))

##select randomly chosen rows from feature class and create new featureclass from selection
containing training points
##then switch selection and create new featureclass containing validating points
#10% of pixels are training points
arcpy.SelectLayerByAttribute_management(class_lyr,"NEW_SELECTION","point_ID" IN'+rand_IDS_10)
out_feature_class = os.path.join(output_tr_location,class_split+"_10_tr.shp")
arcpy.Dissolve_management(class_lyr, out_feature_class,"GRIDCODE")
arcpy.SelectLayerByAttribute_management(class_lyr,"SWITCH_SELECTION")
out_feature_class = os.path.join(output_va_location,class_split+"_10_va.shp")
arcpy.Dissolve_management(class_lyr, out_feature_class,"GRIDCODE")
```

Appendix II.B Random pixels at boundaries sampling design, an example for 10% of total amount of all pixels in a class (source: own Python IDLE output).

```
##RANDOM POINTS BOUNDARY, % AMOUNT
##create boundary featureclass
arcpy.SelectLayerByLocation_management(class_lyr, "WITHIN_A_DISTANCE", boundary, 1,
"NEW_SELECTION")
boundary_pixels = location[:5]+class_split+"_boundary.shp"
arcpy.CopyFeatures_management(class_lyr,boundary_pixels)

#get IDs of pixels contained in featureclass and append them to a list
boundary_l = []
cursor = arcpy.SearchCursor(boundary_pixels)
for row in cursor:
    ID = row.getValue("point_ID")
    boundary_l.append(ID)
del cursor

#randomly choose 10% of rows from IDs in the list
number = count*0.1
try:
    randbound_IDS_10 = sorted(random.sample(boundary_l,int(round(number))))
    randbound_IDS_10 = str(tuple(randbound_IDS_10))
except:
    randbound_IDS_10 = sorted(boundary_l)
    randbound_IDS_10 = str(tuple(randbound_IDS_10))
    print "Except, 10% boundary!"

##select randomly chosen rows from feature class and create new featureclass from selection
containing training points
##then switch selection and create new featureclass containing validating points
#10% of pixels are training points
arcpy.SelectLayerByAttribute_management(class_lyr,"NEW_SELECTION","point_ID"
IN'+randbound_IDS_10)
out_feature_class = os.path.join(output_tr_location,class_split+"_10_bound_tr.shp")
arcpy.Dissolve_management(class_lyr, out_feature_class,"GRIDCODE")
arcpy.SelectLayerByAttribute_management(class_lyr,"SWITCH_SELECTION")
out_feature_class = os.path.join(output_va_location,class_split+"_10_bound_va.shp")
arcpy.Dissolve_management(class_lyr, out_feature_class,"GRIDCODE")
```

Appendix II.C. Clustered sampling design, an example for 5% of total amount of all pixels in a class (source: own Python IDLE output)

```
##RANDOM CLUSTERS, % AMOUNT
##create selection of pixels all of which are at minimum distance or further from given
boundary
#create boundary line
dissolve_cl = location[:5]+class_split+"_diss.shp"
arcpy.Dissolve_management(classl, dissolve_cl, "GRIDCODE")
```



```

boundary = location[:5]+class_split+"_bound.shp"
arcpy.PolygonToLine_management(dissolve_cl, boundary, "IDENTIFY_NEIGHBORS")

#selection for 8-pixel neighbourhood
eight_bour_IDS = []
arcpy.SelectLayerByLocation_management(class_lyr, "WITHIN_A_DISTANCE", boundary, 0,
"NEW_SELECTION")
arcpy.SelectLayerByLocation_management(class_lyr, "WITHIN_A_DISTANCE", boundary, 0,
"SWITCH_SELECTION")
eight_bour_area = location[:5]+class_split+"_8_bour_a.shp"
arcpy.CopyFeatures_management(class_lyr,eight_bour_area)
cursor = arcpy.SearchCursor(eight_bour_area)
for row in cursor:
    ID = row.getValue("point_ID")
    eight_bour_IDS.append(ID)
del cursor

##select central pixels for clusters, amount of points equals 1% of total count of
rows(pixels) of featureclass;
##select n-pixel neighbourhood around these central pixels, round up for classes with count
less than 100
if count > 0 and count < 100:
    number_central = math.ceil(count * 0.01)
else:
    number_central = count * 0.01

#selection of central pixels for 8-pixel neighbourhood
#try and except statement in order to avoid "Sample larger than population" error, in case of
classes containing few pixels
eight_bour_lyr = location[:5]+class_split+"_8_bour_a.lyr"
arcpy.MakeFeatureLayer_management(eight_bour_area, eight_bour_lyr)
try:
    rand_IDS_8_01 = sorted(random.sample(eight_bour_IDS,int(round(number_central))))
    if len(rand_IDS_8_01)> 1:
        rand_IDS_8_01 = str(tuple(rand_IDS_8_01))
        arcpy.SelectLayerByAttribute_management(eight_bour_lyr,"NEW_SELECTION","point_ID"
IN'+rand_IDS_8_01)
    else:
        rand_IDS_8_01 = rand_IDS_8_01[0]
        arcpy.SelectLayerByAttribute_management(eight_bour_lyr,"NEW_SELECTION","point_ID"
=''+str(rand_IDS_8_01))
except:
    print "Except central pixels 8 neighbourhood!"
    rand_IDS_8_01 = sorted(eight_bour_IDS)
    if len(rand_IDS_8_01)> 1:
        rand_IDS_8_01 = str(tuple(rand_IDS_8_01))
        arcpy.SelectLayerByAttribute_management(eight_bour_lyr,"NEW_SELECTION","point_ID"
IN'+rand_IDS_8_01)
    else:
        rand_IDS_8_01 = rand_IDS_8_01[0]
        arcpy.SelectLayerByAttribute_management(eight_bour_lyr,"NEW_SELECTION","point_ID"
=''+str(rand_IDS_8_01))

#selection of 8-pixel neighbourhood around central pixels
arcpy.SelectLayerByLocation_management(class_lyr,"WITHIN_A_DISTANCE", eight_bour_lyr,
0,"NEW_SELECTION")
eight_bour = location[:5]+class_split+"_8_bour.shp"
arcpy.CopyFeatures_management(class_lyr,eight_bour)

#get IDs of pixels of 8-pixel neighbourhood and append them to a list
rand_IDS_8_clusters = []
cursor = arcpy.SearchCursor(eight_bour)
for row in cursor:
    ID = row.getValue("point_ID")
    rand_IDS_8_clusters.append(ID)
del cursor

##select randomly 5% of IDs from rand IDs 8 clusters
number = count * 0.05
try:
    rand_IDS_8_05 = sorted(random.sample(rand_IDS_8_clusters,int(round(number))))
    rand_IDS_8_05 = str(tuple(rand_IDS_8_05))
except:
    print "Except, 5% IDs clusters!"
    rand_IDS_8_05 = sorted(rand_IDS_8_clusters)
    rand_IDS_8_05 = str(tuple(rand_IDS_8_05))

```

```

##select randomly chosen IDs and create new featureclass from selection containing training
points
##then switch selection and create new featureclass containing validating points
#5% of pixels are training points
arcpy.SelectLayerByAttribute_management(class_lyr,"NEW_SELECTION","point_ID"
IN'+rand_IDS_8_05)
out_feature_class = os.path.join(output_tr_location,class_split+"_05_cl_tr.shp")
arcpy.Dissolve_management(class_lyr, out_feature_class,"GRIDCODE")
arcpy.SelectLayerByAttribute_management(class_lyr,"SWITCH_SELECTION")
out_feature_class = os.path.join(output_va_location,class_split+"_05_cl_va.shp")
arcpy.Dissolve_management(class_lyr, out_feature_class,"GRIDCODE")

```

Appendix II.D. “Real life scenario” sampling design, an example for 10% of total amount of pixels in a class (source: own Python IDLE output).

```

## REAL LIFE SCENARIO, % AMOUNT
#create lyr for seed point
seed_point = os.path.join(folder_seedpoint, location+"_"+class_split+"_seed.shp" )
seed_point_lyr = class_split+"_"+feature_split[:5]+"_seedpoint.lyr"
arcpy.MakeFeatureLayer_management (seed_point, seed_point_lyr)

#assign final count 10%
final_count = int(round(count * 0.10))
arcpy.SelectLayerByLocation_management (class_lyr, "CONTAINS", seed_point_lyr, "", "NEW_SELECTION")
pix_count = 1

#grow region
while pix_count < final_count:

arcpy.SelectLayerByLocation_management (class_lyr, "WITHIN_A_DISTANCE", class_lyr, "0", "ADD_TO_SELECTION")
    curr_count = arcpy.GetCount_management (class_lyr)
    pix_count = int(curr_count.getOutput(0))

number = pix_count - final_count

#create sub shapefiles, or in case there are no pixels exceeding, save as final
training/validating pixels shapefile
if number == 0:
    out_feature_class_tr = os.path.join(output_tr_location,class_split+"_10_area_tr.shp")
    arcpy.Dissolve_management (class_lyr, out_feature_class_tr,"GRIDCODE")
    arcpy.SelectLayerByAttribute_management (class_lyr,"SWITCH_SELECTION")
    out_feature_class_va = os.path.join(output_va_location,class_split+"_10_area_va.shp")
    arcpy.Dissolve_management (class_lyr, out_feature_class_va,"GRIDCODE")
else:
    #temp training pixels
    out_feature_class_tr = feature_split+"_"+class_split+"_10_area_temp_tr.shp"
    arcpy.CopyFeatures_management (class_lyr, out_feature_class_tr)

    #dissolve and create boundary from sub shapefiles
    out_feature_diss = feature_split+"_"+class_split+"_10_area_diss.shp"
    arcpy.Dissolve_management (class_lyr, out_feature_diss)
    out_feature_class = feature_split+"_"+class_split+"_10_area_boundary.shp"
    arcpy.PolygonToLine_management(out_feature_diss,out_feature_class)

    #temp validating pixels
    arcpy.SelectLayerByAttribute_management(class_lyr,"SWITCH_SELECTION")
    out_feature_class_va = feature_split+"_"+class_split+"_10_area_temp_va.shp"
    arcpy.Dissolve_management (class_lyr, out_feature_class_va,"GRIDCODE")

    #define pixels on the boundary, and append their ID to a list
    boundary_lyr = feature_split+"_"+class_split+"_10_area_boundary.lyr"
    arcpy.MakeFeatureLayer_management (out_feature_class,boundary_lyr)
    out_feature_class_tr_lyr = feature_split+"_"+class_split+"_10_area_temp_tr.lyr"
    arcpy.MakeFeatureLayer_management (out_feature_class_tr,out_feature_class_tr_lyr)

arcpy.SelectLayerByLocation_management(out_feature_class_tr_lyr,"WITHIN_A_DISTANCE",boundary_lyr,"0","NEW_SELECTION")
    pixels_to_remove = feature_split+"_"+class_split+"_10_pixels.shp"
    arcpy.CopyFeatures_management (out_feature_class_tr_lyr, pixels_to_remove)

```

```
cursor = arcpy.SearchCursor(pixels_to_remove)
pixels = []
for row in cursor:
    ID = row.getValue("point_ID")
    pixels.append(ID)
del cursor

#randomly choose from IDs the amount which exceeds the final count
pix_to_remove = sorted(random.sample(pixels,int(round(number))))
pix_to_remove = str(tuple(pix_to_remove))

#select pixels which are extra and add them to validating pixels
arcpy.SelectLayerByAttribute_management(out_feature_class_tr_lyr,"NEW_SELECTION","point_ID"
IN'+pix_to_remove)
out_feature_class_va2 = feature_split+" "+class_split+"_10_area_temp2_va.shp"
arcpy.Dissolve_management(out_feature_class_tr_lyr, out_feature_class_va2,"GRIDCODE")
out_feature_class_va_final =os.path.join(output_va_location,class_split+"_10_area_va.shp")
arcpy.Merge_management ([out_feature_class_va,out_feature_class_va2],
out_feature_class_va_final)

#select training pixels
arcpy.SelectLayerByAttribute_management(out_feature_class_tr_lyr,"SWITCH_SELECTION")
out_feature_class_tr = os.path.join(output_tr_location,class_split+"_10_area_tr.shp")
arcpy.Dissolve_management(out_feature_class_tr_lyr, out_feature_class_tr,"GRIDCODE")
```

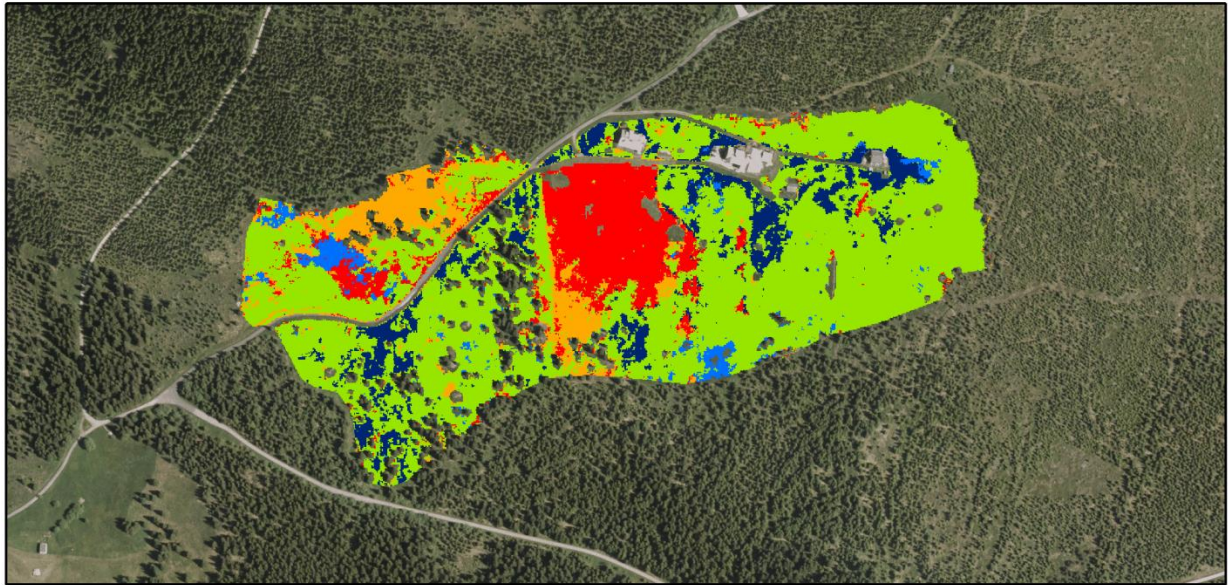
Appendix III – Classification maps of Přední Rennerovky

Appendix III.A. Classification map of Přední Rennerovky for the SVM classifier (upper).

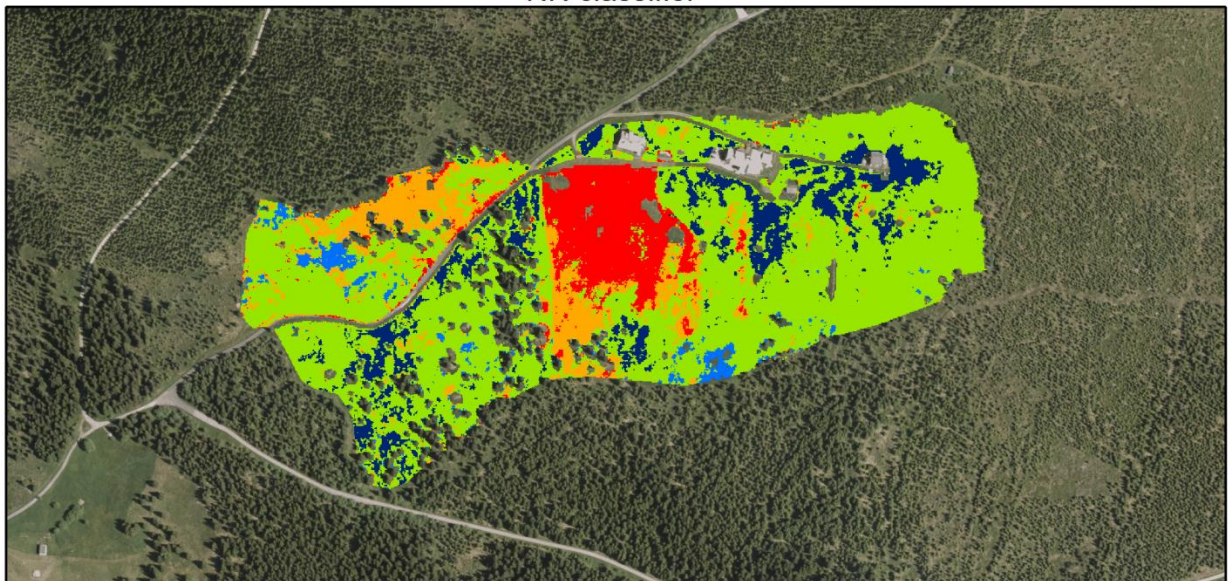
Appendix III.B. Classification map of Přední Rennerovky for the NN classifier (lower).

(source: own ArcGIS output; background of the maps: orthophoto (ČÚZK, orthophoto)).

SVM classifier

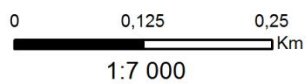


NN classifier



Classes

	Oligotrophic grasslands		Degraded meadows dominated by Dicotyledons
	Mesophile grasslands		Degraded meadows dominated by grass species
	Waterlogged grasslands		



Appendix IV – Error matrices for the classification of Přední Rennerovky

Appendix IV.A. The error matrix for the SVM classifier. Numbers of classes: 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pixels / %	2	3	4	5	6	Total (pixels)	UA	CE
2	4866 / 80.03%	498 / 4.08%	39 / 1.00%	38 / 0.59%	550 / 1.63%	5991	4866 / 5991 81.22%	1125 / 5991 18.78%
3	316 / 5.20%	10474 / 85.90%	223 / 5.73%	99 / 1.53%	498 / 1.48%	11610	10474 / 11610 90.22%	1136 / 11610 9.78%
4	37 / 0.61%	261 / 2.14%	1753 / 45.05%	74 / 1.14%	320 / 0.95%	2445	1753 / 2445 71.70%	692 / 2445 28.30%
5	0 / 0.00%	7 / 0.06%	87 / 2.24%	3629 / 55.93%	900 / 2.67%	4623	3629 / 4623 78.50%	994 / 4623 21.50%
6	861 / 14.16%	953 / 7.82%	1789 / 45.98%	2648 / 40.81%	31389 / 93.26%	37640	31389 / 37640 83.39%	6251 / 37640 16.61%
Total (pixels)	6080	12193	3891	6488	33657	62309		
PA	4866 / 6080 80.03%	10474 / 12193 85.90%	1753 / 3891 45.05%	3629 / 6488 55.93%	31389 / 33657 93.26%		OA = 83.63%	
OE	1214 / 6080 19.97%	1719 / 12193 14.10%	2138 / 3891 54.95%	2859 / 6488 44.07%	2268 / 33659 6.74%		kappa = 0.74	

Appendix IV.B. The error matrix for the NN classifier (source: own ENVI analysis).

Pixels / %	2	3	4	5	6	Total (pixels)	UA	CE
2	4902 / 80.74%	1129 / 9.26%	155 / 3.98%	196 / 3.02%	928 / 2.76%	7310	4902 / 7310 7.06%	2408 / 7310 32.94%
3	231 / 3.80%	9014 / 73.92%	28 / 0.72%	32 / 0.49%	110 / 0.33%	9415	9014 / 9415 95.74%	401 / 9415 4.26%
4	107 / 1.76%	230 / 1.89%	1462 / 37.57%	4 / 0.06%	269 / 0.80%	2072	1462 / 2072 70.56%	610 / 2072 29.44%
5	0 / 0.00%	3 / 0.02%	226 / 5.81%	3647 / 56.21%	1303 / 3.87%	5179	3647 / 5179 70.42%	1532 / 5179 29.58%
6	831 / 13.69%	1818 / 14.91%	2020 / 51.91%	2609 / 40.21%	31031 / 92.24%	38309	31031 / 38309 81.00%	7278 / 38309 19.00%
Total (pixels)	6071	12194	3891	6488	33641	62285		
PA	4902 / 6071 80.74%	9014 / 12194 73.92%	1462 / 3891 37.57%	3647 / 6488 56.21%	31031 / 33641 92.24%		OA = 80.37%	
OE	1169 / 6071 19.26%	3180 / 12194 26.08%	2429 / 3891 62.43%	2841 / 6488 43.79%	2610 / 33641 7.76%		kappa = 0.68	

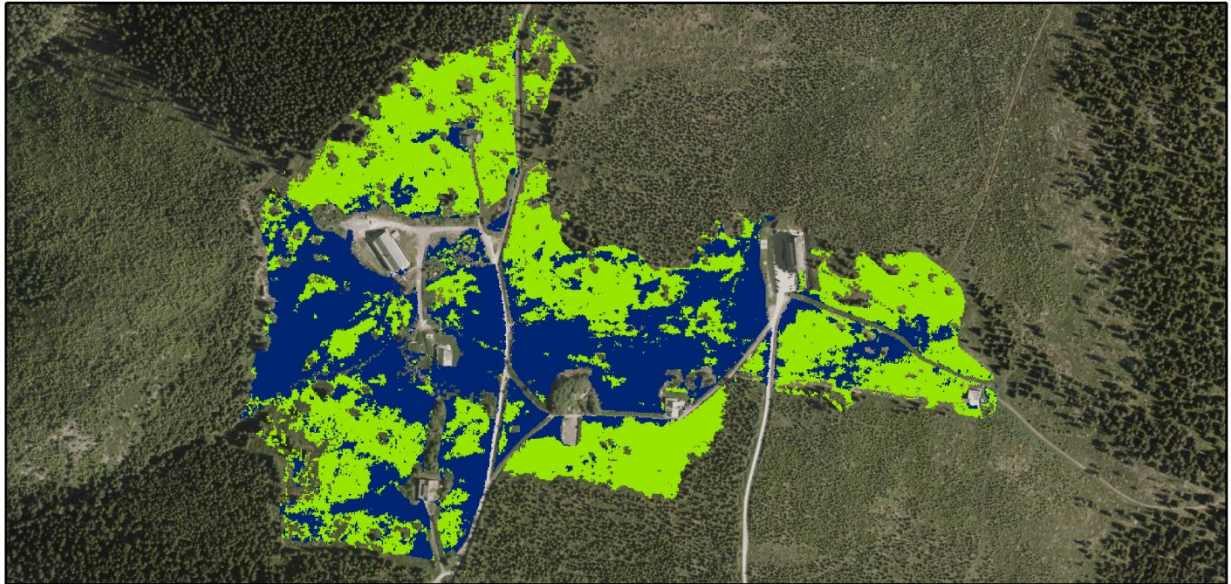
Appendix V – Classification maps of Friesovy Boudy

Appendix V.A. Classification map of Friesovy Boudy for the SVM classifier (upper).

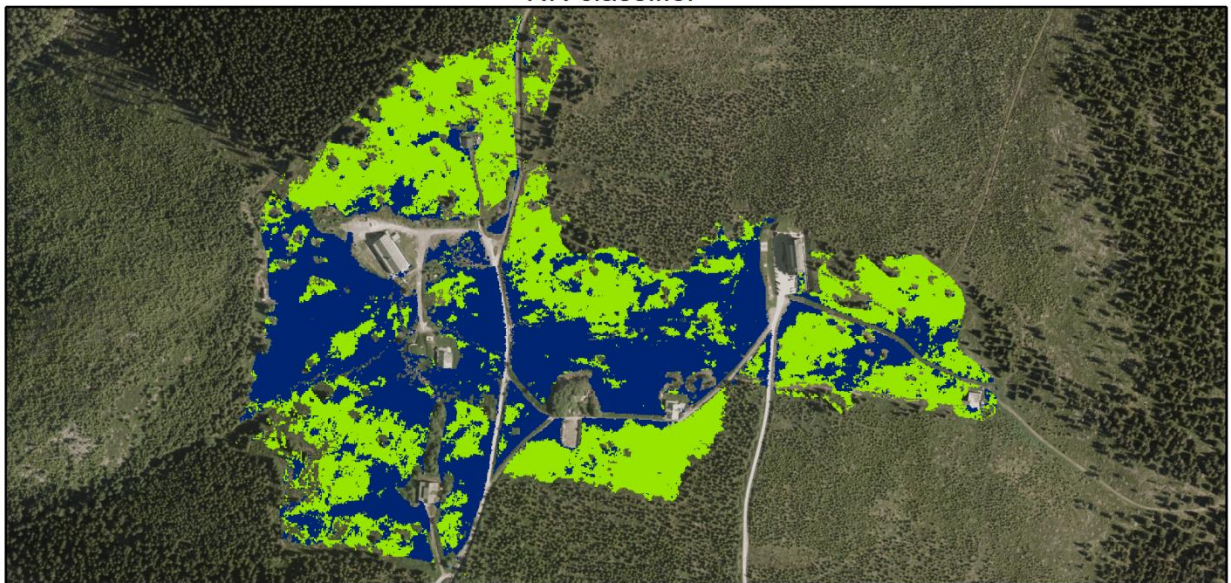
Appendix V.B. Classification map of Friesovy Boudy for the NN classifier (lower).

(source: own ArcGIS output; background of the maps: orthophoto (ČÚZK, orthophoto)).



SVM classifier



NN classifier



Classes

-  Degraded meadows dominated by Dicotyledons
-  Degraded meadows dominated by grass species

0 0,125 0,25
Km
1:7 000



Appendix VI– Error matrices for the classification of Friesovy Boudy

Appendix VI.A. The error matrix for the SVM classifier. Numbers of classes: 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species. Abbreviations: OA – overall accuracy UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pixels / %	4	5	6	7	Total (pixels)	UA	CE
4	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0	0 / 0 0.00%	0 / 0 0.00%
5	56 / 14.00%	27367 / 83.50%	3359 / 7.93%	0 / 0.00%	30782	27367 / 30782 88.91%	3415 / 30782 11.09%
6	344 / 86.00%	5409 / 16.50%	39015 / 92.07%	202 / 100.00%	44970	39015 / 44970 86.76%	5955 / 44970 13.24%
7	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0	0 / 0 0.00%	0 / 0 0.00%
Total (pixels)	400	32776	42374	202	75752		
PA	0 / 400 0.00%	27367 / 32776 83.50%	39015 / 42374 92.07%	0 / 202 0.00%		OA = 87.63%	
OE	400 / 400 100%	5409 / 32776 16.50%	3359 / 42374 7.93%	202 / 202 100%		kappa = 0.75	

Appendix VI.B. The error matrix for the NN classifier (source: own ENVI analysis).

Pixels / %	4	5	6	7	Total (pixels)	UA	CE
4	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0	0 / 0 0.00%	0 / 0 0.00%
5	62 / 15.50%	28712 / 87.58%	4914 / 11.59%	0 / 0.00%	33688	28712 / 33688 85.23%	4976 / 33688 14.77%
6	338 / 84.50%	4071 / 12.42%	37496 / 88.41%	202 / 100.00%	42107	37496 / 42107 89.05%	4611 / 42107 10.95%
7	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0	0 / 0 0.00%	0 / 0 0.00%
Total (pixels)	400	32783	42410	202	75795		
PA	0 / 400 0.00%	28712 / 32783 87.58%	37496 / 42410 88.41%	0 / 202 0.00%		OA = 87.35%	
OE	400 / 400 100%	4071 / 32783 12.42%	4914 / 42410 11.59%	202 / 202 100%		kappa = 0.75	

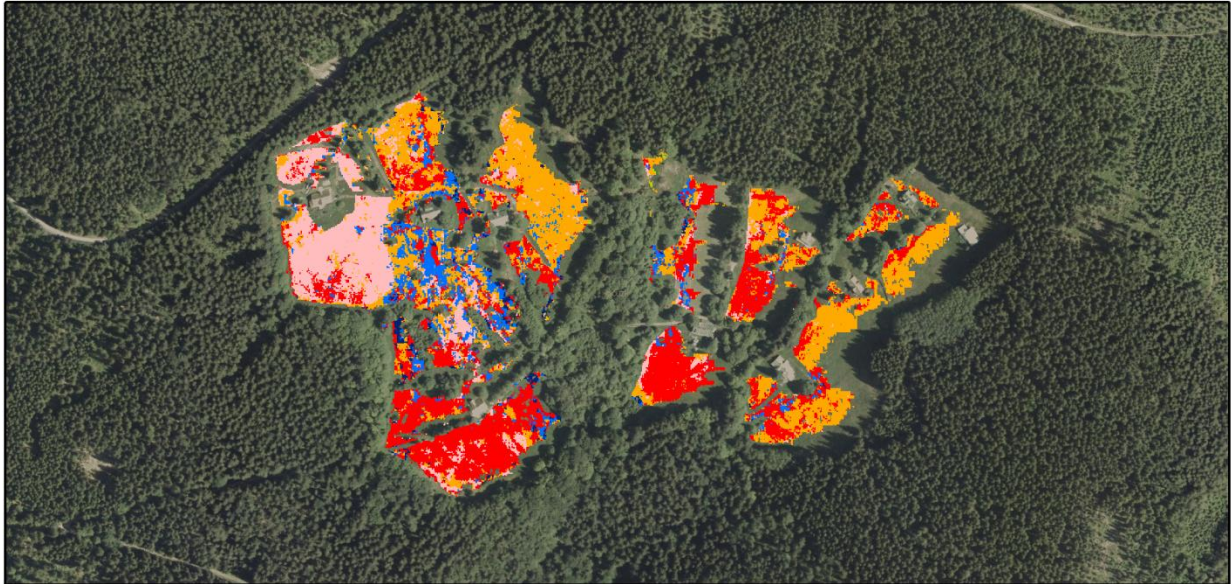
Appendix VII – Classification maps of Husí Boudy

Appendix VII.A. Classification map of Friesovy Boudy for the SVM classifier (upper).

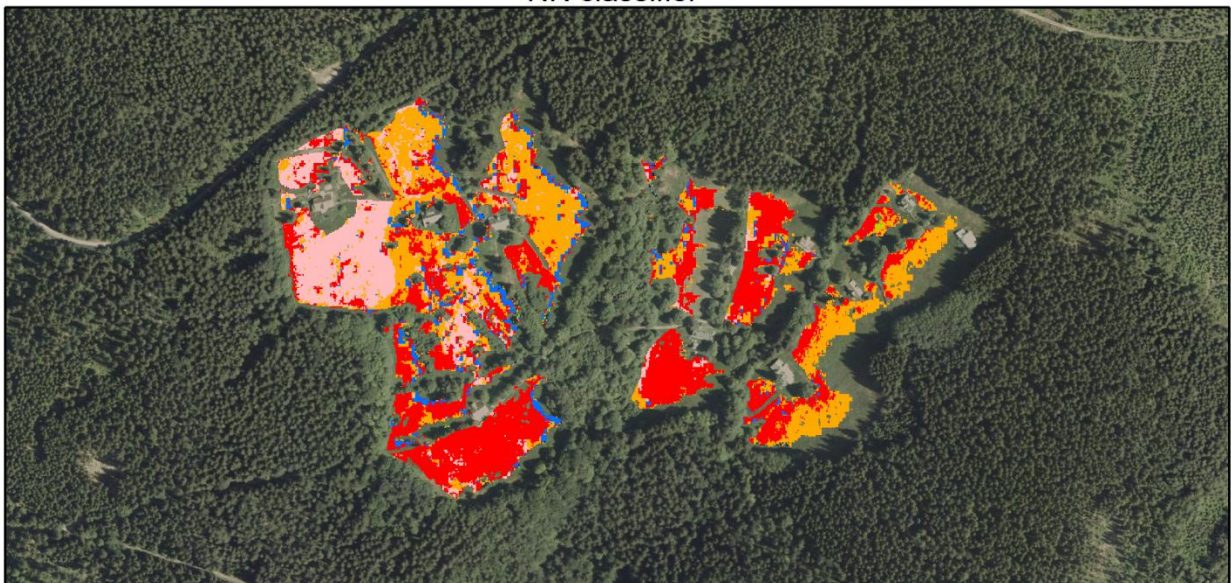
Appendix VII.B. Classification map of Friesovy Boudy for the NN classifier (lower).

(source: own ArcGIS output; background of the maps: orthophoto (ČÚZK, orthophoto)).

SVM classifier

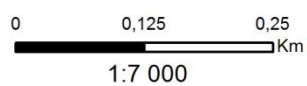


NN classifier



Classes

	Grasslands dominated by <i>Nardus stricta</i>		Waterlogged grasslands
	Oligotrophic grasslands		Degraded meadows dominated by Dicotyledons
	Mesophile grasslands		Degraded meadows dominated by grass species



Appendix VIII - Error matrices for the classification of Husí Boudy

Appendix VIII.A. The error matrix for the SVM classifier. Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	Tot(pix)	UA	CE
1	7351 / 66.38%	880 / 6.65%	1112 / 7.46%	1026 / 14.38%	133 / 4.39%	20 / 11.17%	10522	7351 / 10522 69.86%	3171 / 10522 30.14%
2	1638 / 14.79%	9724 / 73.44%	2410 / 16.16%	2238 / 31.37%	1000 / 32.98%	115 / 64.25%	17125	9724 / 17125 56.78%	7401 / 17125 43.22%
3	1580 / 14.27%	1932 / 14.59%	10666 / 71.53%	1483 / 20.79%	993 / 32.75%	19 / 10.61%	16673	10666 / 16673 63.97%	6007 / 16673 36.03%
4	476 / 4.30%	579 / 4.37%	507 / 3.40%	2213 / 31.02%	540 / 17.81%	10 / 5.59%	4325	2213 / 4325 51.17%	2112 / 4325 48.83%
5	28 / 0.25%	119 / 0.90%	214 / 1.44%	170 / 2.38%	366 / 12.07%	1 / 0.56%	898	366 / 898 40.76%	532 / 898 59.24%
6	1 / 0.01%	7 / 0.05%	2 / 0.01 %	4 / 0.06%	0 / 0.00%	14 / 7.82%	28	14 / 28 50.00%	14 / 28 50.00%
Tot (pix)	11074	13241	14911	7134	3032	179	49571		
PA	7351 / 11074 66.38%	9724 / 13241 73.44%	10666 / 14911 71.53%	2213 / 7134 31.02%	366 / 3032 12.07%	14 / 179 7.82%		OA = 61.20%	
OE	3723 / 11074 33.62%	3517 / 13241 26.56%	4245 / 14911 28.47%	4921 / 7134 68.98%	2666 / 3032 87.93%	165 / 179 92.18%		kappa = 0.48	

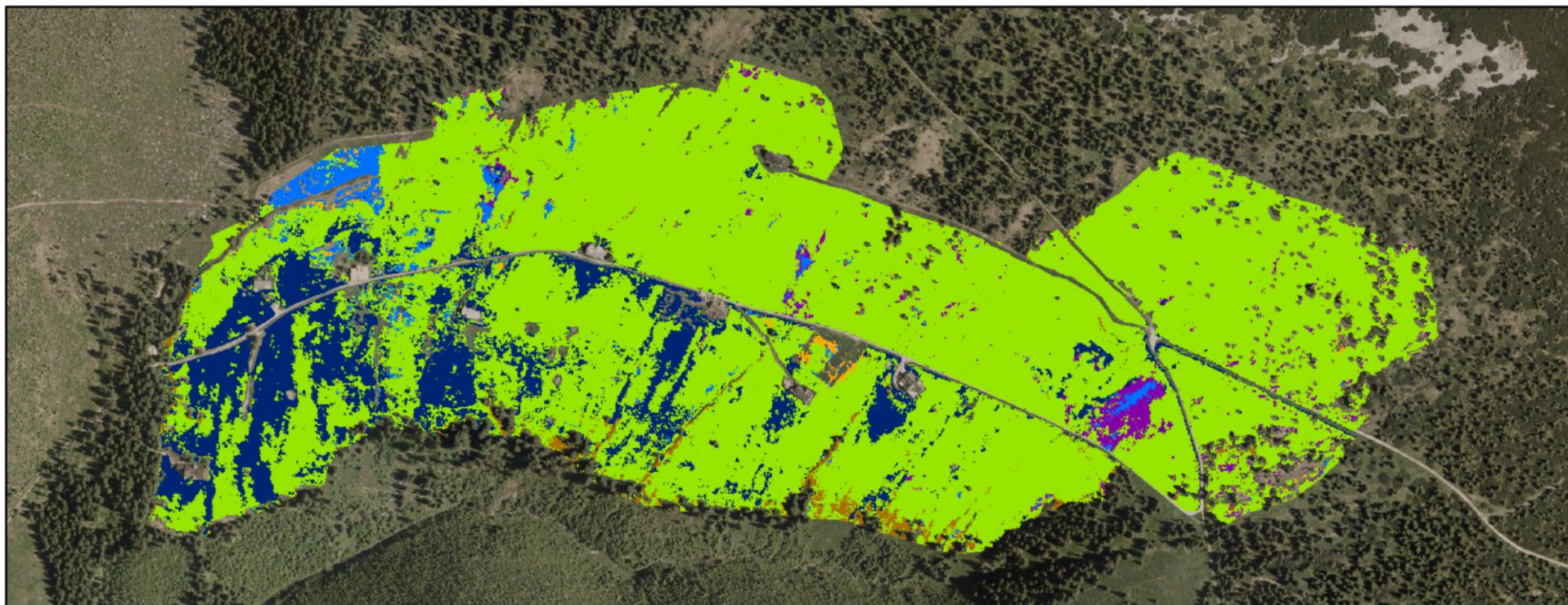
Appendix VIII.B. The error matrix for the NN classifier (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	Tot(pix)	UA	CE
1	7251 / 65.47%	934 / 7.04 %	508 / 3.41%	1042 / 14.61%	128 / 4.21%	17 / 9.50%	9880	7251 / 9880 73.39%	2629 / 9880 26.61%
2	1616 / 14.59%	8774 / 66.18%	1432 / 9.60%	2808 / 39.36%	961 / 31.64%	15 / 8.38%	15606	8774 / 15606 56.22%	6832 / 15606 43.78%
3	1939 / 17.51%	2720 / 20.52%	12298 / 82.45%	2411 / 33.80%	1292 / 42.54%	100 / 55.87%	20760	12298 / 20760 59.24%	8462 / 20760 40.76%
4	233 / 2.10%	788 / 5.94%	657 / 4.40%	853 / 11.96%	637 / 20.97%	28 / 15.64%	3196	853 / 3196 26.69%	2343 / 3196 73.31%
5	10 / 0.09%	25 / 0.19%	18 / 0.12%	14 / 0.20%	13 / 0.43%	15 / 8.38%	95	13 / 95 13.68%	82 / 95 86.32%
6	27 / 0.24%	17 / 0.13%	2 / 0.01%	6 / 0.08%	6 / 0.20%	4 / 2.23%	62	4 / 62 6.45%	58 / 62 93.55%
Tot (pix)	11076	13258	14915	7134	3037	179	49599		
PA	7251 / 11076 65.47%	8774 / 13258 66.18%	12298 / 14915 82.45%	853 / 7134 11.96%	13 / 3037 0.43%	4 / 179 2.23%		OA = 58.86%	
OE	3825 / 11076 34.53%	4484 / 13258 33.82%	2617 / 14915 17.55%	6281 / 7134 88.04%	3024 / 3037 99.57%	175 / 179 97.77%		kappa = 0.44	


Appendix IX – Classification maps for Klínové Boudy

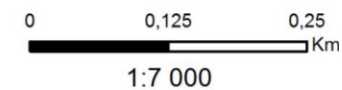
Appendix IX.A. Classification map of Klínové Boudy for the SVM classifier (source: own ArcGIS output; background of the map: orthophoto (ČÚZK, orthophoto)).

SVM classifier



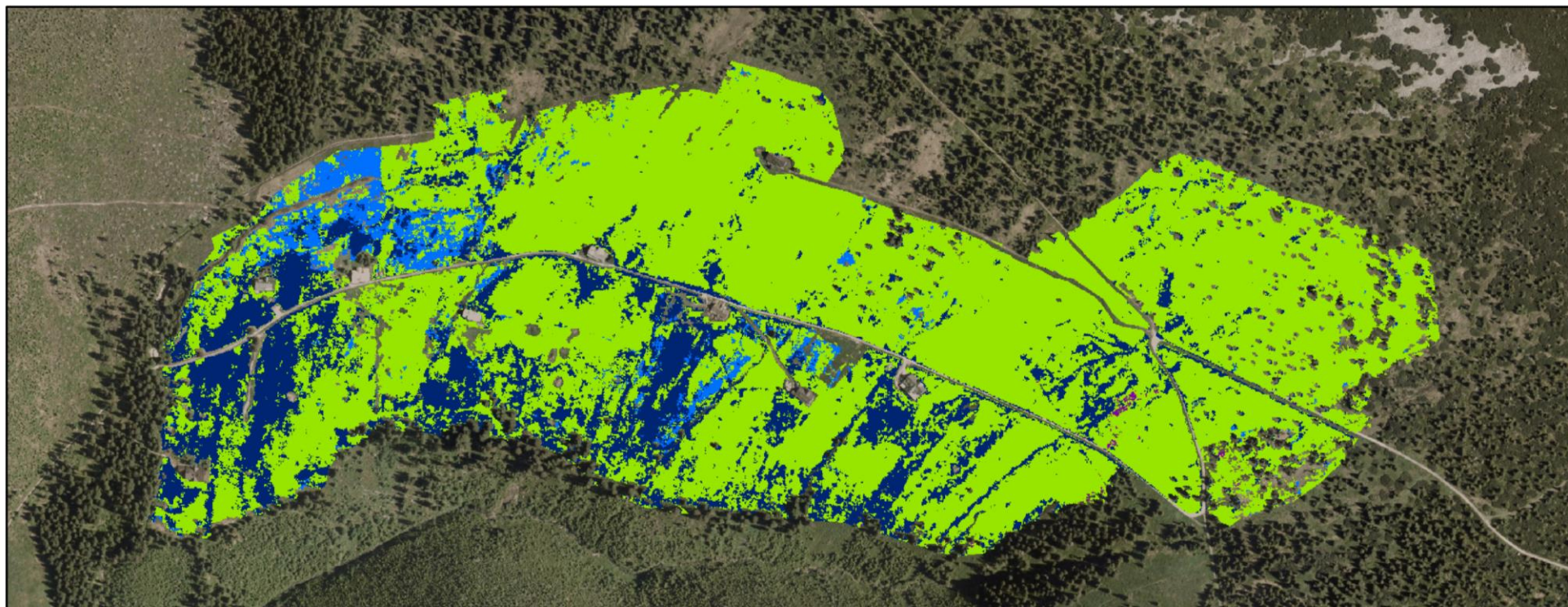
Classes

 Oligotrophic grasslands	 Stands dominated by <i>Vaccinium</i> species
 Waterlogged grasslands	 Vegetation of springs
 Degraded meadows dominated by Dicotyledons	 Tall-fern vegetation
 Degraded meadows dominated by grass species	








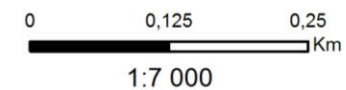
Appendix IX.B. Classification map of Klínové Boudy for the NN classifier (source: own ArcGIS output; background of the map: orthophoto (ČÚZK, orthophoto)).

NN classifier



Classes

- | | |
|---|--|
|  Waterlogged grasslands |  Stands dominated by <i>Vaccinium</i> species |
|  Degraded meadows dominated by Dicotyledons |  Vegetation of springs |
|  Degraded meadows dominated by grass species | |



Appendix X - Error matrices for the classification of Klínové Boudy

Appendix X.A. The error matrix for the SVM classifier. Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species, 8 – vegetation of springs; 9 – tall-fern vegetation. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pix / %	1	2	4	5	6	7	8	9	Tot (pix)	UA	CE
1	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
2	4 0,09%	585 16,85%	84 0,45%	30 0,05%	219 0,11%	0 0,00%	65 0,37%	0 0,00%	987	585 / 987 59,27%	402 / 987 40,73%
4	2 0,05%	167 4,81%	3708 20,03%	399 44,95%	1268 0,66%	49 0,37%	621 3,54%	0 0,00%	6214	3708 / 6214 59,67%	2506 / 6214 40,33%
5	437 10,01%	323 9,31%	2030 10,97%	27656 44,95%	11594 6,01%	1036 7,82%	896 5,10%	66 28,33%	44038	27656 / 44038 62,80%	16382 / 44038 37,20%
6	3788 87,80%	2360 67,99%	11773 63,60%	33448 52,74%	177938 92,20%	10316 77,85%	13840 78,83%	165 70,82%	252628	177938 / 252628 70,43%	74690 / 252628 29,57%
7	73 1,67%	2 0,06%	65 0,35%	462 075%	734 0,38%	1734 13,09%	69 0,39%	1 0,43%	3140	1734 / 3140 55,22%	1406 / 3140 44,78%
8	60 1,37%	34 0,98%	850 4,59%	525 0,85%	1247 0,65%	116 0,88%	2066 11,77%	0 0,00%	4898	2066 / 4898 42,18%	2832 / 4898 57,82%
9	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	1 0,43%	1	1 / 1 100,00%	0 / 1 0,00%
Tot (pix)	4364	3471	18510	61520	193000	13251	17557	233	311906		
PA	0 / 4364 0,00%	585 / 3471 16,85%	3708 / 18510 20,03%	27656 / 61520 44,95%	177938 / 193000 92,20%	1734 / 13251 13,09%	2066 / 17557 11,77%	1 / 233 0,43%		OA = 68,51%	
OE	4364 / 4364 100,00%	2886 / 3471 83,15%	14802 / 18510 79,97%	33864 / 61520 55,05%	15062 / 193000 7,80%	11517 / 13251 86,91%	15491 / 17557 88,23%	232 / 233 99,57%		kappa = 0.33	

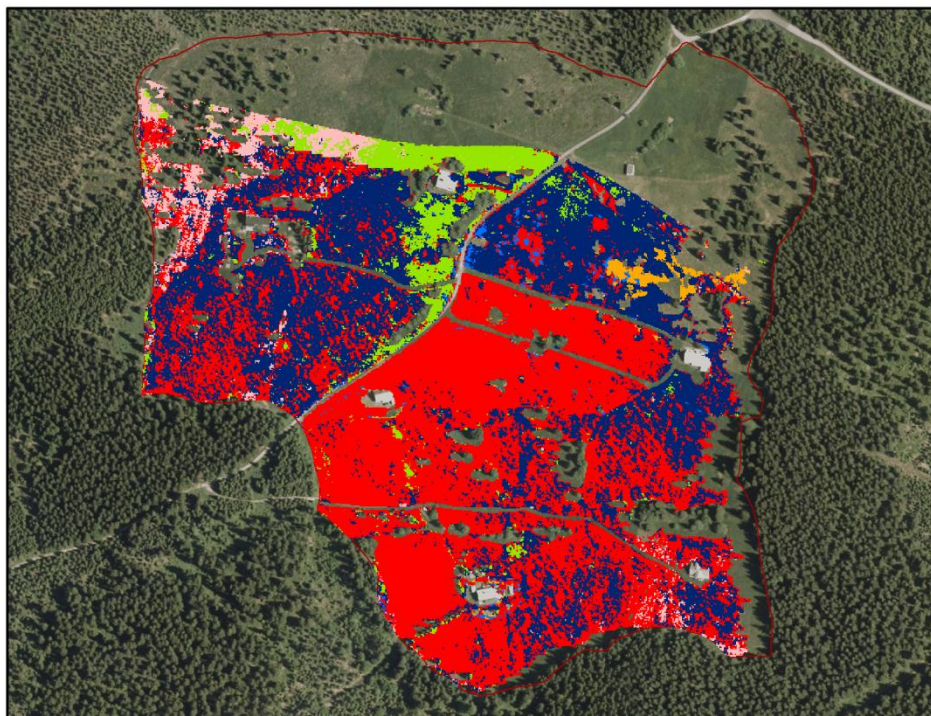
Appendix X.B. The error matrix for the NN classifier. Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species, 8 – vegetation of springs; 9 – tall-fern vegetation. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pix / %	1	2	4	5	6	7	8	9	Tot (pix)	UA	CE
1	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
2	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
4	70 1,60%	775 22,33%	5881 31,84%	1987 3,23%	2782 1,44%	61 0,46%	866 4,97%	0 0,00%	12422	5881 / 12422 47,34%	6541 / 12422 52,66%
5	1410 32,29%	636 18,32%	3493 18,91%	29960 48,70%	19689 10,21%	4243 32,28%	1887 10,82%	97 41,63%	61415	29960 / 61415 48,78%	31455 / 61415 51,22%
6	2882 65,99%	2060 59,35%	8974 48,59%	29545 48,02%	170191 88,30%	8749 66,56%	14607 83,75%	136 58,37%	237145	170191 / 237145 71,77%	66154 / 237145 28,23%
7	0 0,00%	0 0,00%	1 0,01%	12 0,02%	14 0,01%	19 0,14%	0 0,00%	0 0,00%	46	19 / 46 41,30%	27 / 46 58,70%
8	5 0,11%	0 0,00%	121 0,66%	15 0,02%	71 0,04%	72 0,55%	81 0,46%	0 0,00%	365	81 / 365 22,19%	284 / 365 77,81%
9	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
Tot (pix)	4367	3471	18470	61519	192747	13144	17442	233	311393		
PA	0 / 4367 0,00%	0 / 3471 0,00%	5881 / 18470 31,84%	29960 / 61519 48,70%	170191 / 192747 88,30%	19 / 13144 0,14%	81 / 17442 0,46%	0 / 233 0,00%		OA = 66,20%	
OE	4367 / 4367 100,0%	3471 / 3471 100,00%	12589 / 18470 68,16%	31559 / 61519 51,30%	22556 / 192747 11,70%	13125 / 13144 99,86%	17361 / 17442 99,54%	233 / 233 100,00%		kappa = 0.31	

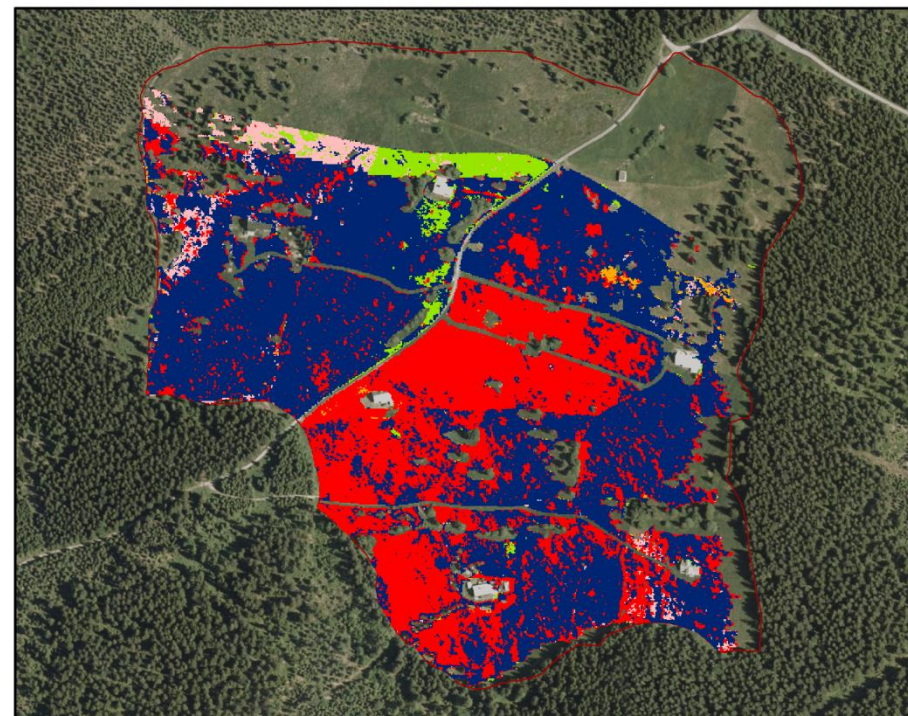
Appendix XI – Classification maps for Lahrový Boudy on the flight line 08_43

Appendix XI.A. Classification map of Lahrový Boudy on the flight line 08_43 for the SVM classifier (left). **Appendix XI.B.** Classification map of Lahrový Boudy on the flight line 08_43 for the NN classifier (right) (source: own ArcGIS output; background of the maps: orthophoto (ČÚZK, orthophoto)).

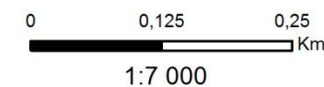
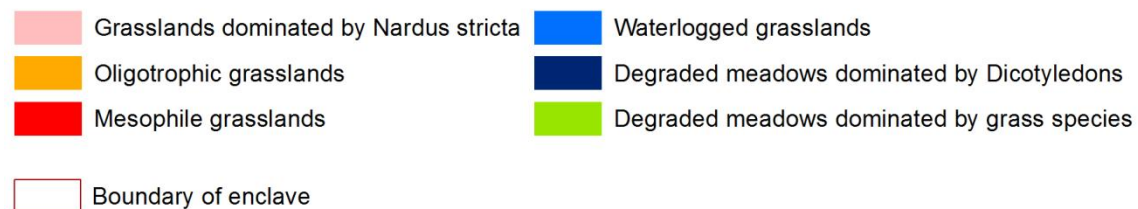
SVM classifier



NN classifier



Classes



Appendix XII – Error matrices for the classification of Lahrový Boudy on the flight line 08_43

Appendix XII.A. The error matrix for the SVM classifier. Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	7	Tot(pix)	UA	CE
1	2061 / 41.62%	62 / 1.40%	714 / 1.13%	4 / 0.13%	340 / 0.83%	550 / 3.96%	7 / 8.75%	3738	2061 / 3738 55.14%	1677 / 3738 44.86%
2	3 / 0.06%	962 / 21.65%	61 / 0.10%	15 / 0.50%	329 / 0.80%	19 / 0.14%	0 / 0.00%	1389	962 / 1389 69.26%	427 / 1389 30.74%
3	1407 / 28.41%	1455 / 32.74%	50498 / 79.74%	588 / 19.63%	12519 / 30.55%	1968 / 14.16%	17 / 21.25%	68452	50498 / 68452 73.77%	17954 / 68452 26.23%
4	8 / 0.16%	60 / 1.35%	79 / 0.12%	316 / 10.55%	121 / 0.30%	26 / 0.19%	0 / 0.00%	610	316 / 610 51.80%	294 / 610 48.20%
5	1202 / 24.27%	1890 / 42.53%	11363 / 17.94%	2069 / 69.08%	26790 / 65.37%	6587 / 47.41%	55 / 68.75%	49956	26790 / 49956 53.63%	23166 / 49956 46.37%
6	271 / 5.47%	15 / 0.34%	614 / 0.97%	3 / 0.10%	880 / 2.15%	4745 / 34.15%	1 / 1.25%	6529	4745 / 6529 72.68%	1784 / 6529 27.32%
7	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0	0 / 0 0.00%	0 / 0 0.00%
Tot (pix)	4952	4444	63329	2995	40979	13895	80	130674		
PA	2061 / 4952 41.62%	962 / 4444 21.65%	50498 / 63329 79.74%	316 / 2995 10.55%	26790 / 40979 65.37%	4745 / 13895 34.15%	0 / 80 0.00%		OA = 65.33%	
OE	2891 / 4952 58.38%	3482 / 4444 78.35%	12831 / 63329 20.26%	2679 / 2995 89.45%	14189 / 40979 34.63%	9150 / 13895 65.85%	80 / 80 100.00%		kappa = 0.44	

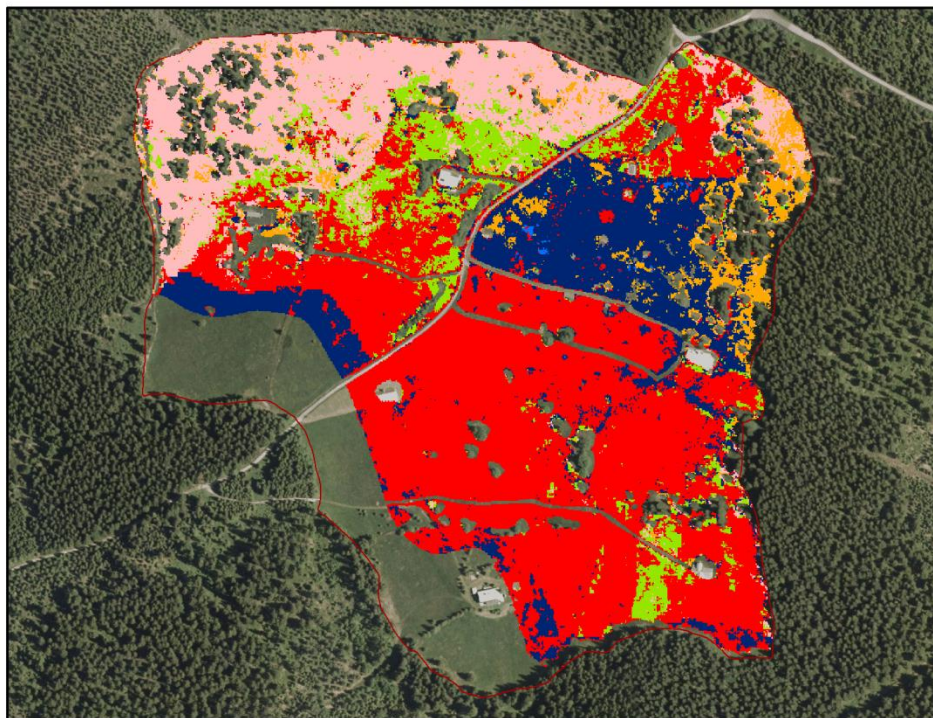
Appendix XII.B. The error matrix for the NN classifier (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	7	Tot(pix)	UA	CE
1	1543 / 31.17%	142 / 3.20%	571 / 0.90%	0 / 0.00%	163 / 0.40%	480 / 3.46%	1 / 1.25%	2900	1543 / 2900 53.21%	1357 / 2900 46.79%
2	33 / 0.67%	462 / 10.40%	60 / 0.09%	0 / 0.00%	65 / 0.16%	19 / 0.14%	0 / 0.00%	639	462 / 639 72.30%	177 / 639 27.70%
3	681 / 13.76%	491 / 11.05%	37560 / 59.31%	694 / 23.17%	3734 / 9.18%	966 / 6.96%	5 / 6.25%	44131	37560 / 44131 85.11%	6571 / 44131 14.89%
4	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0	0 / 0 0.00%	0 / 0 0.00%
5	2554 / 51.60%	3336 / 75.07%	25023 / 39.51%	2300 / 76.79%	36942 / 90.15%	9470 / 68.23%	74 / 92.50%	79699	36942 / 79699 46.35%	42757 / 79699 53.65%
6	139 / 2.81%	13 / 0.29%	115 / 0.18%	1 / 0.03%	75 / 0.18%	2945 / 21.22%	0 / 0.00%	3288	2945 / 3288 89.57%	343 / 3288 10.43%
7	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0 / 0.00%	0	0 / 0 0.00%	0 / 0 0.00%
Tot (pix)	4950	4444	63329	2995	40979	13880	80	130657		
PA	1543 / 4950 31.17%	462 / 4444 10.40%	37560 / 63329 59.31%	0 / 2995 0.00%	36942 / 40979 90.15%	2945 / 13880 21.22%	0 / 80 0.00%		OA = 60.81%	
OE	3407 / 4950 68.83%	3982 / 4444 89.60%	25769 / 63329 40.69%	2995 / 2995 100.00%	4037 / 40979 9.85%	10935 / 13880 78.78%	80 / 80 100.00%		kappa = 0.39	

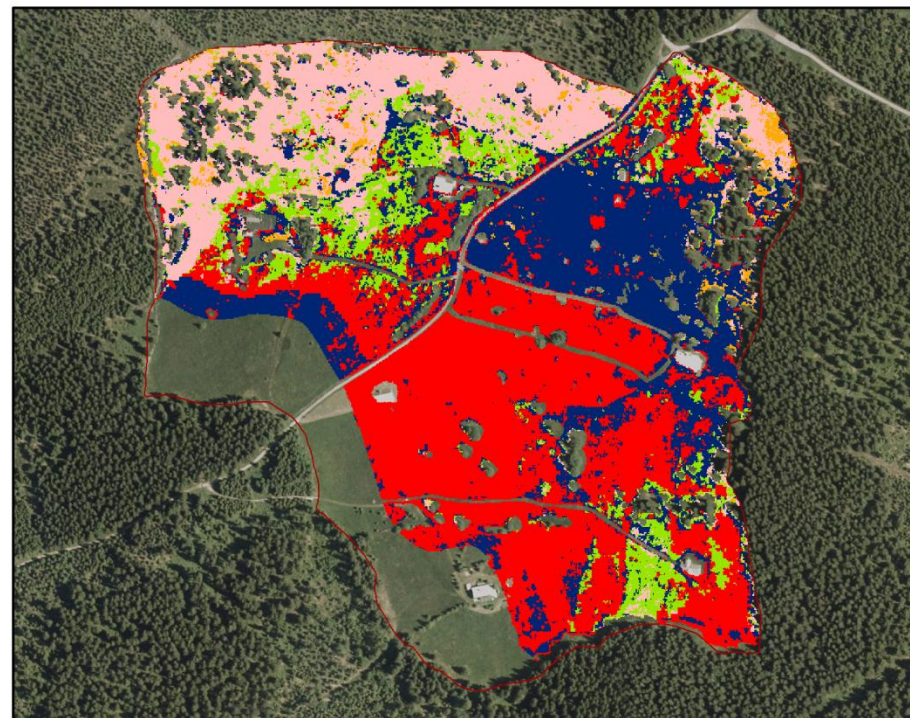
Appendix XIII – The classification maps for Lahrový Boudy on the flight line 10_45

Appendix XIII.A. Classification map of Lahrový Boudy on the flight line 10_45 for the SVM classifier (left). **Appendix XIII.B.** Classification map of Lahrový Boudy on the flight line 10_45 for the NN classifier (right) (source: own Arc

SVM classifier



NN classifier

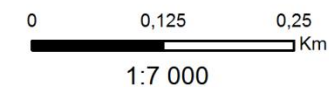


Classes

- | | |
|---|---|
|  Grasslands dominated by <i>Nardus stricta</i> |  Waterlogged grasslands |
|  Oligotrophic grasslands |  Degraded meadows dominated by Dicotyledons |
|  Mesophile grasslands |  Degraded meadows dominated by grass species |

 Boundary of enclave

GIS output; background of the maps: orthophoto (ČÚZK, orthophoto)).



Appendix XIV - Error matrices for the classification of Lahrový Boudy on the flight line 10_45

Appendix XIV.A. The error matrix for the SVM classifier. Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species, 8 – vegetation of springs. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	7	8	Tot(pix)	UA	CE
1	14039 / 84,25%	1143 / 13,00%	974 / 1,58%	3 / 0,11%	539 / 1,61%	2680 / 15,85%	77 / 63,64%	0 / 0,00%	19455	14039 / 19455 72,16%	5416 / 19455 27,48%
2	609 / 3,65%	2779 / 31,60%	125 / 0,20%	443 / 15,59%	834 / 2,49%	677 / 4,00%	0 / 0,00%	1 / 1,20%	5468	2779 / 5468 50,82%	2689 / 5468 49,18%
3	814 / 4,88%	1322 / 15,03%	53171 / 86,01%	166 / 5,84%	15518 / 46,24%	6588 / 38,96%	7 / 5,79%	32 / 38,55%	77618	53171 / 77618 68,50%	24447 / 77618 31,50%
4	3 / 0,02%	20 / 0,23%	9 / 0,01%	84 / 2,96%	72 / 0,21%	2 / 0,01%	0 / 0,00%	0 / 0,00%	190	84 / 190 44,21%	106 / 190 55,79%
5	88 / 0,53%	3429 / 38,99%	3623 / 5,86%	2133 / 75,05%	15639 / 46,60%	531 / 3,14%	0 / 0,00%	50 / 60,24%	25493	15639 / 25493 61,35%	9854 / 25493 38,65%
6	1111 / 6,67%	101 / 1,15%	3916 / 6,33%	13 / 0,46%	955 / 2,85%	6432 / 38,04%	37 / 30,58%	0 / 0,00%	12565	6432 / 12565 51,19%	6133 / 12565 48,81%
7	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
8	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
Tot (pix)	16664	8794	61818	2842	33557	16910	121	83	140789		
PA	14039 / 16664 84,25%	2779 / 8794 31,60%	53171 / 61818 86,01%	84 / 2842 2,96%	15639 / 33557 46,60%	6432 / 16910 38,04%	0 / 121 0,00%	0 / 83 0,00%		OA = 65,45%	
OE	2625 / 16664 15,75%	6015 / 8794 68,40%	8647 / 61818 13,99%	2758 / 2842 97,04%	17918 / 33557 53,40%	10478 / 16910 61,96%	121 / 121 100,00%	83 / 83 100,00%		kappa = 0.50	

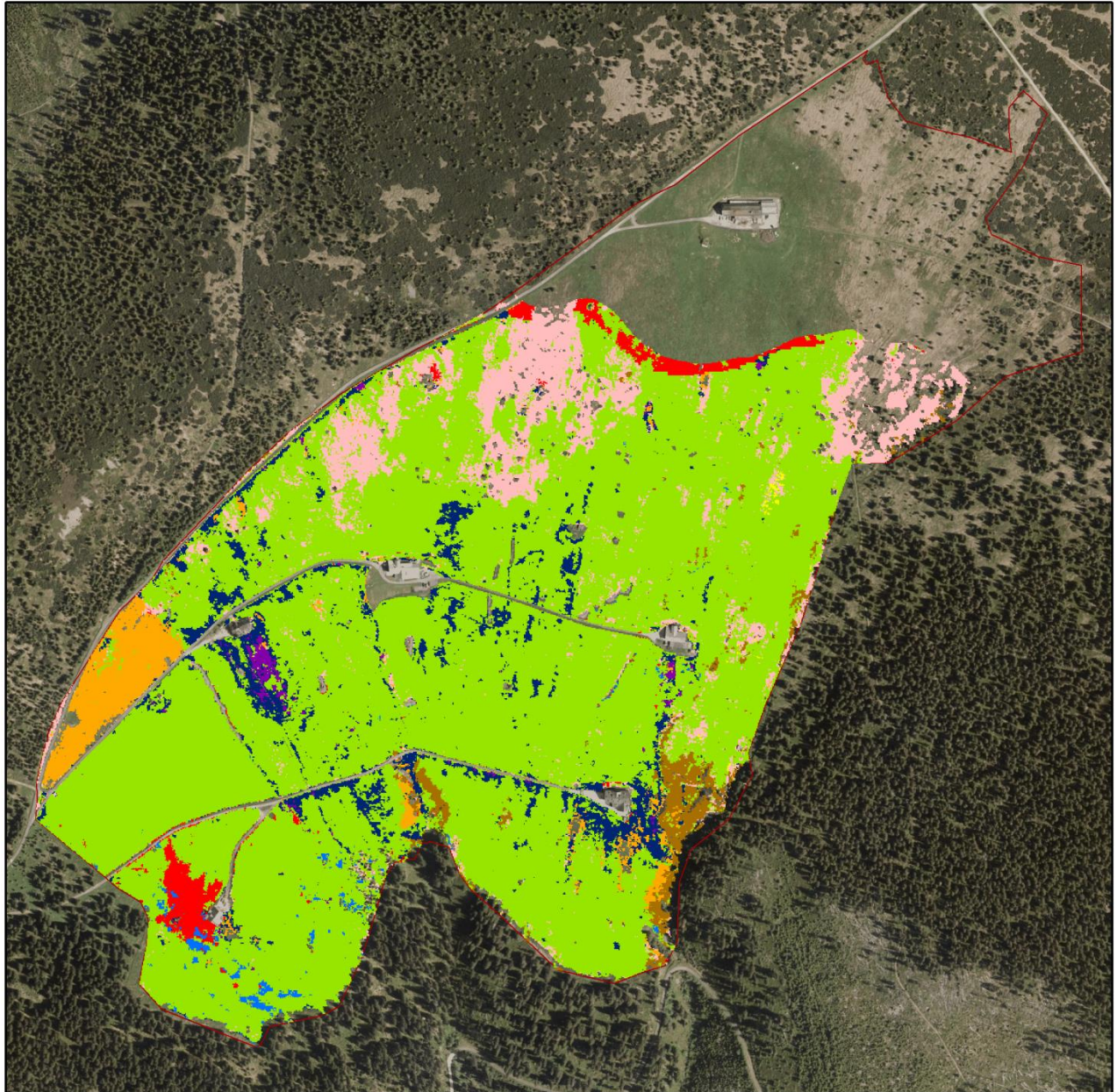
Appendix XIV.B. The error matrix for the NN classifier (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	7	8	Tot(pix)	UA	CE
1	13205 / 79,24%	1368 / 15,56%	1482 / 2,40%	19 / 0,67%	634 / 1,89%	3052 / 18,05%	65 / 53,71%	0 / 0,00%	19825	13205 / 19825 66,61%	6620 / 19825 33,39%
2	642 / 3,85%	1046 / 11,89%	11 / 0,02%	61 / 2,15%	170 / 0,51%	422 / 2,50%	1 / 0,83%	0 / 0,00%	2353	1046 / 2353 44,45%	1307 / 2353 55,55%
3	317 / 1,90%	1521 / 17,30%	46738 / 75,61%	342 / 12,03%	11651 / 34,72%	2786 / 16,48%	1 / 0,83%	33 / 39,76%	63389	46738 / 63389 73,73%	16651 / 63389 26,27%
4	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
5	1094 / 6,57%	4777 / 54,32%	8052 / 13,03%	2420 / 85,15%	19054 / 56,78%	4136 / 24,46%	12 / 9,92%	50 / 60,24%	39595	19054 / 39595 48,12%	20541 / 39595 51,88%
6	1406 / 80,44%	82 / 0,93%	5535 / 8,95%	0 / 0,00%	2048 / 6,10%	6514 / 38,55%	42 / 34,71%	0 / 0,00%	15627	6514 / 15627 41,68%	9113 / 15627 58,32%
7	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
8	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
Tot (pix)	16664	8794	61818	2842	33557	16910	121	83	140789		
PA	13205 / 16664 79,24%	1046 / 8794 11,89%	46738 / 61818 75,61%	0 / 2842 0,00%	19054 / 33557 56,78%	6514 / 16910 38,52%	0 / 121 0,00%	0 / 83 0,00%		OA = 61,48%	
OE	3459 / 16664 20,76%	7748 / 8794 88,11%	15080 / 61818 24,39%	2842 / 2842 100,00%	14503 / 33557 43,22%	10396 / 16910 61,48%	121 / 121 100,00%	83 / 83 100,00%		kappa = 0.45	











Appendix XV – Classification maps for Zadní Rennerovky on the flight line 10_58

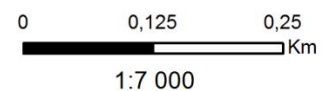
Appendix XV.A. Classification map of Zadní Rennerovky on the flight line 10_58 for the SVM classifier
(source: own ArcGIS output; background of the map: orthophoto (ČÚZK, orthophoto)).

SVM classifier



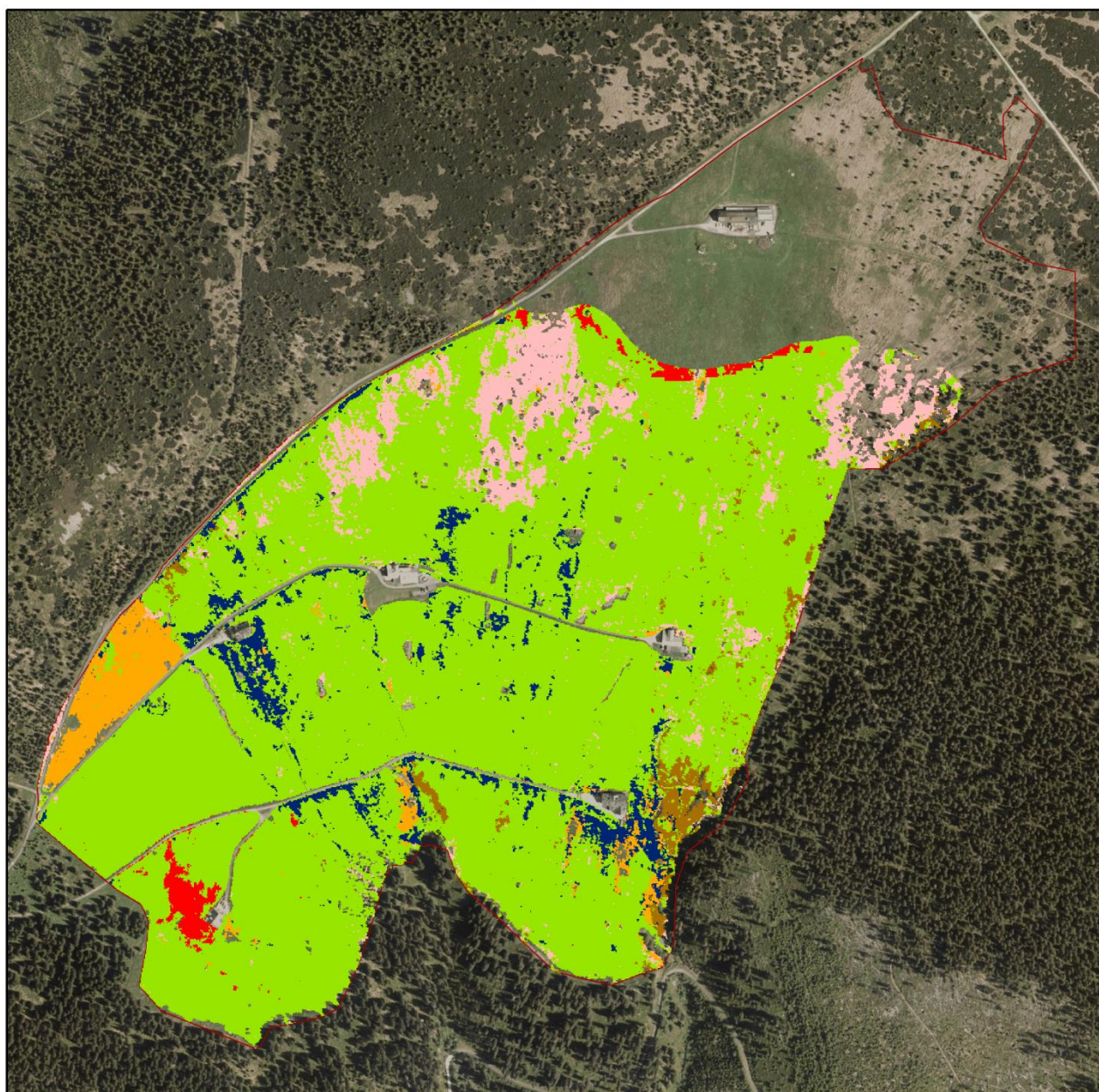
Classes

	Grasslands dominated by <i>Nardus stricta</i>		Degraded meadows dominated by grass species
	Oligotrophic grasslands		Stands dominated by <i>Vaccinium</i> species
	Mesophile grasslands		Vegetation of springs
	Waterlogged grasslands		Tall-fern vegetation
	Degraded meadows dominated by Dicotyledons		
	Boundary of enclave		










Appendix XV.B. Classification map of Zadní Rennerovky on the flight line 10_58 for the NN classifier (source: own ArcGIS output; background of the map: orthophoto (ČÚZK, orthophoto)).

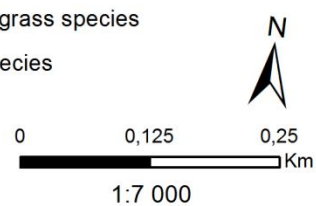
NN classifier



Classes

- | | |
|---|--|
|  Grasslands dominated by <i>Nardus stricta</i> |  Degraded meadows dominated by Dicotyledons |
|  Oligotrophic grasslands |  Degraded meadows dominated by grass species |
|  Mesophile grasslands |  Stands dominated by <i>Vaccinium</i> species |

 Boundary of enclave



Appendix XVI – Error matrices for the classification of Zadní Rennerovky on the flight line 10_58

Appendix XVI.A. The error matrix for the SVM classifier. Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species; 8 – vegetation of springs; 9 – tall-fern vegetation. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	7	8	9	Tot(pix)	UA	CE
1	17763 49.41%	757 5.47%	2123 10.63%	83 2.21%	1215 1.87%	2774 1.95%	474 5.95%	34 0.72%	379 16.73%	25602	17763 / 25602 69.38%	7839 / 25602 30.62%
2	86 0.24%	9159 66.17%	18 0.09%	4 0.18%	817 1.26%	825 0.58%	104 1.31%	29 0.62%	0 0.00%	11042	9159 / 11042 82.95%	1883 / 11042 17.05%
3	327 0.91%	69 0.50%	3302 16.54%	66 1.75%	219 0.34%	865 0.61%	5 0.06%	2 0.04%	0 0.00%	4855	3302 / 4855 68.01%	1553 / 4855 31.99%
4	2 0.01%	0 0.00%	22 0.11%	691 18.37%	134 0.21%	272 0.19%	3 0.04%	32 0.68%	0 0.00%	1156	691 / 1156 59.78%	465 / 1156 40.22%
5	440 1.22%	942 6.81%	179 0.90%	23 0.61%	8043 12.39%	4329 3.05%	191 2.40%	836 17.81%	0 0.00%	14983	8043 / 14983 53.68%	6940 / 14983 46.32%
6	16967 47.19%	2835 20.48%	14248 71.35%	2894 76.93%	53984 83.19%	132425 93.17%	4143 52.04%	2861 60.94%	1863 82.25%	232220	132425 / 232220 57.03%	99795 / 232220 42.97%
7	323 0.90%	12 0.09%	19 0.10%	1 0.03%	230 0.35%	500 0.35%	3031 38.07%	0 0.00%	11 0.49%	4127	3031 / 4127 73.44%	1096 / 4127 26.56%
8	42 0.12%	67 0.48%	50 0.25%	0 0.00%	244 0.38%	114 0.08%	4 0.05%	901 19.19%	0 0.00%	1422	901 / 1422 63.36%	521 / 1422 36.64%
9	1 0.00%	0 0.00%	7 0.04%	0 0.00%	7 0.01%	23 0.02%	6 0.08%	0 0.00%	12 0.53%	56	12 / 56 21.43%	44 / 56 78.57%
Tot (pix)	35951	13841	19968	3762	64893	142127	7961	4695	2265	295463		
PA	17763 / 35951 49.41%	9159 / 13841 66.17%	3302 / 19968 16.54%	691 / 3762 18.37%	8043 / 64893 12.39%	132425 / 142127 93.17%	3031 / 7961 38.07%	901 / 4695 19.19%	12 / 2265 0.53%		OA = 59.34%	
OE	18188 / 35951 50.59%	4682 / 13841 33.83%	16666 / 19968 83.46%	3071 / 3762 81.63%	56850 / 64893 87.61%	9702 / 142127 6.83%	4930 / 7961 61.93%	3794 / 4695 80.81%	2253 / 2265 99.47%		kappa = 0.32	

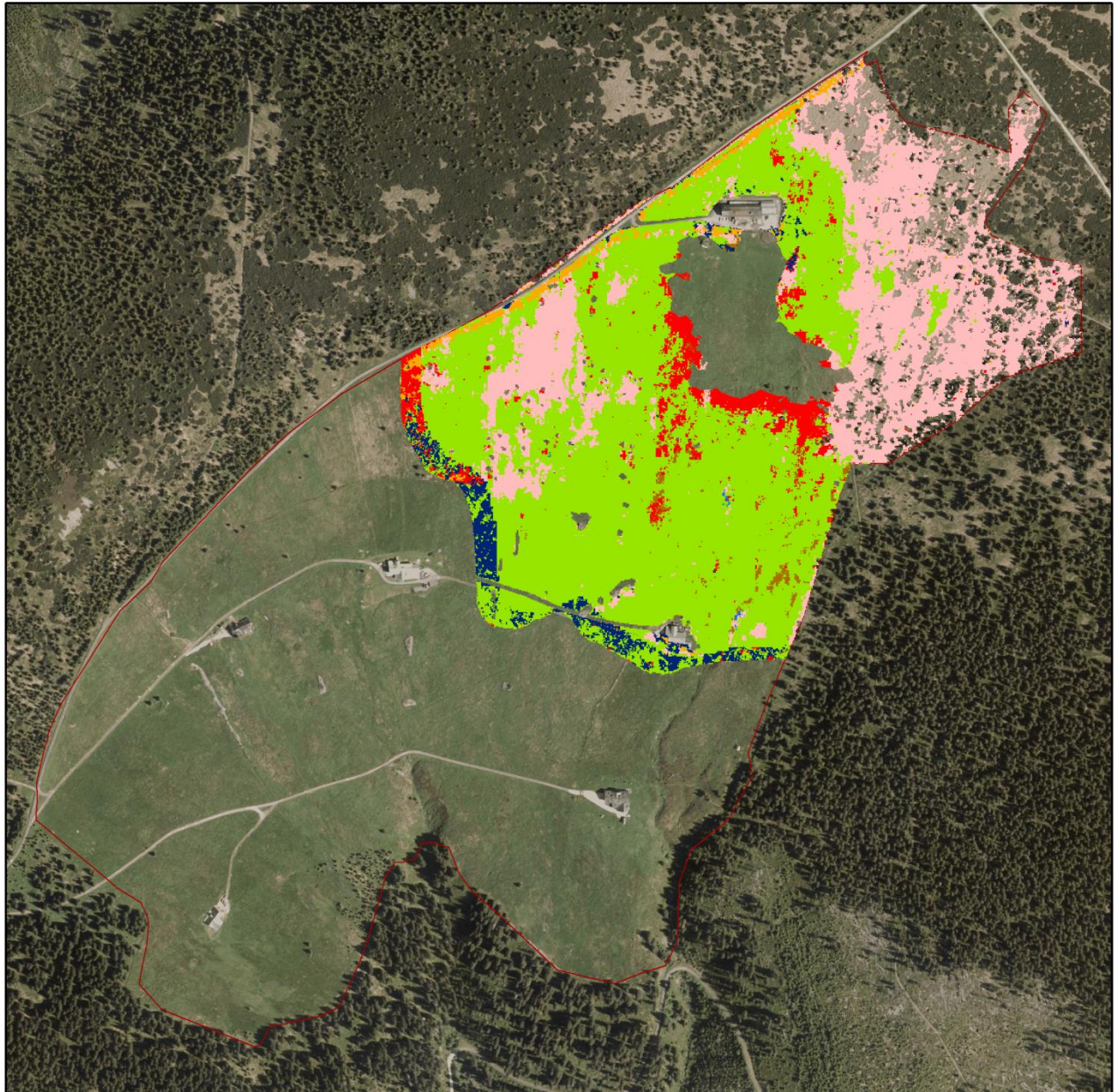
Appendix XVI.B. The error matrix for the NN classifier (source: own ENVI analysis). . Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species, 8 – vegetation of springs; 9 – tall-fern vegetation. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	7	8	9	Tot(pix)	UA	CE
1	15504 43,14%	550 3,97%	1906 9,55%	66 1,75%	879 1,35%	1930 1,36%	361 4,53%	48 1,02%	252 11,13%	21496	15504 / 21496 72,13%	5992 / 21496 27,87%
2	116 0,32%	9025 65,20%	2495 12,49%	5 0,13%	1011 1,56%	946 0,67%	80 1,00%	53 1,13%	0 0,00%	11303	9025 / 11303 79,85%	2278 / 11303 20,15%
3	210 0,58%	31 0,22%	0 0,00%	86 2,29%	124 0,19%	263 1,36%	5 0,06%	0 0,00%	0 0,00%	3214	2495 / 3214 77,63%	719 / 3214 22,37%
4	0 0,00%	0 0,00%	54 0,27%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
5	157 0,44%	744 5,38%	67 0,34%	0 0,00%	5874 9,05%	3072 2,16%	177 2,22%	1215 25,88%	0 0,00%	11293	5874 / 11293 52,01%	5419 / 11293 47,99%
6	19488 54,23%	3379 24,41%	15420 77,22%	3603 95,77%	56604 87,23%	135256 95,17%	4544 57,08%	3379 71,97%	2006 88,57%	243679	135256 / 243679 55,51%	108423 / 243679 44,49%
7	464 1,29%	112 0,81%	26 0,13%	2 0,05%	401 0,62%	660 0,46%	2794 35,10%	0 0,00%	7 0,31%	4466	2794 / 4466 62,56%	1672 / 4466 37,44%
8	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
9	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0	0 / 0 0,00%	0 / 0 0,00%
Tot (pix)	35939	13841	19968	3762	64893	142127	7961	4695	2265	295451		
PA	15504 / 35939 43,14%	9025 / 13841 65,20%	2495 / 19968 12,48%	0 / 3762 0,00%	5874 / 64893 9,05%	135256 / 142127 95,17%	2794 / 7961 35,10%	0 / 4695 0,00%	0 / 2265 0,00%		OA = 57,86%	
OE	20435 / 35939 56,86%	4816 / 13841 34,80%	17473 / 19968 87,51%	3762 / 3762 100,00%	59019 / 64893 90,95%	6871 / 142127 4,83%	5167 / 7961 64,90%	4695 / 4695 100,00%	2265 / 2265 100,00%		kappa = 0.28	








Appendix XVII – The classification maps for Zadní Rennerovky on the flight line 10_20


Appendix XVII.A. Classification map of Zadní Rennerovky on the flight line 10_20 for the SVM classifier (source: own ArcGIS output; background of the map: orthophoto (ČÚZK, orthophoto)).

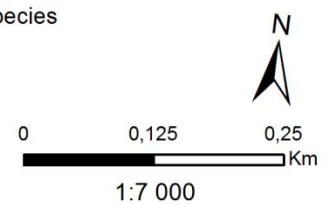
SVM classifier



Classes

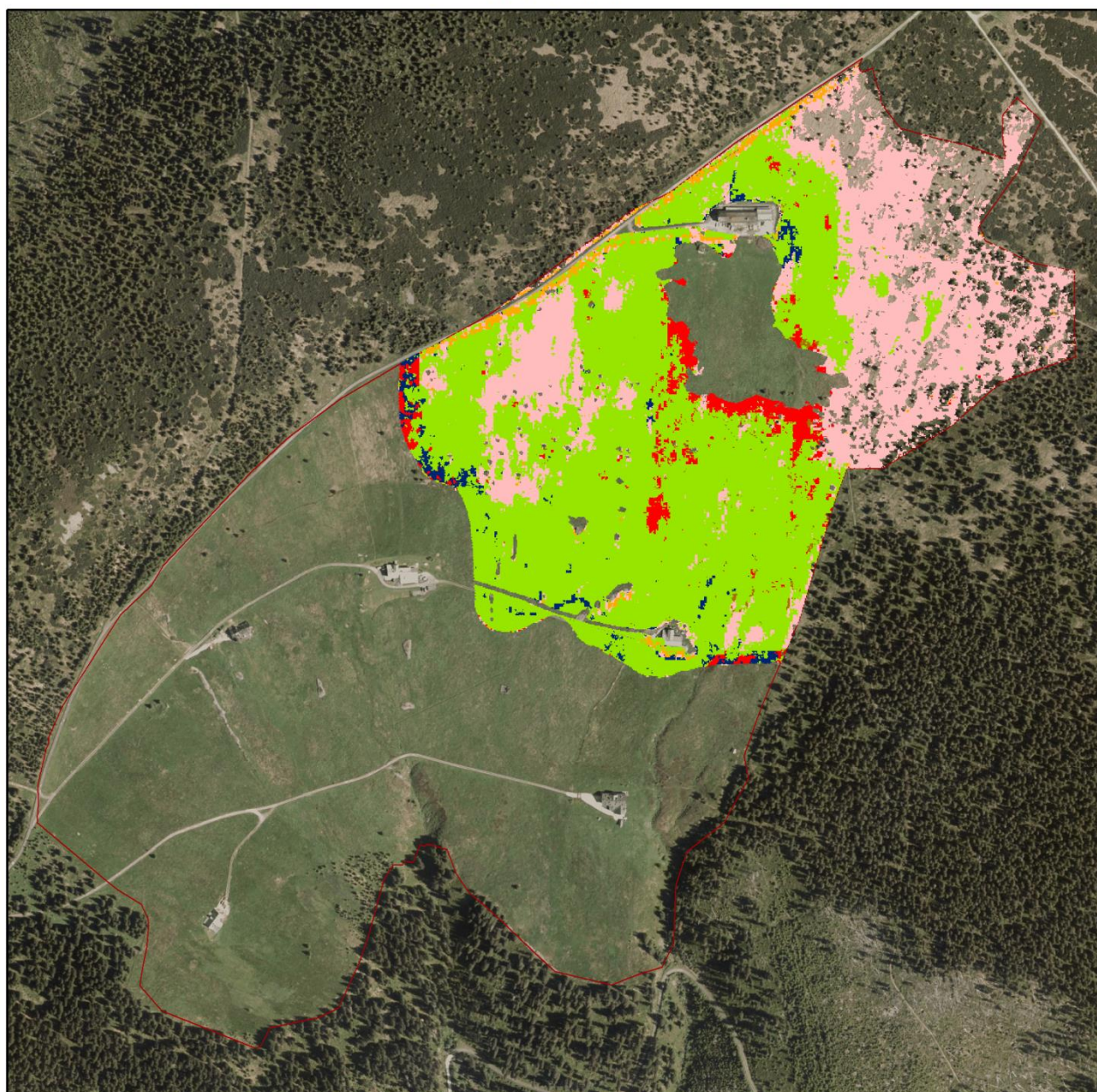
- | | |
|---|--|
|  Grasslands dominated by <i>Nardus stricta</i> |  Degraded meadows dominated by Dicotyledons |
|  Oligotrophic grasslands |  Degraded meadows dominated by grass species |
|  Mesophile grasslands |  Stands dominated by <i>Vaccinium</i> species |
|  Waterlogged grasslands | |

 Boundary of enclave








Appendix XVII.B. Classification map of Zadní Rennerovky on the flight line 10_20 for the NN classifier (source: own ArcGIS output; background of the map: orthophoto (ČÚZK, orthophoto)).

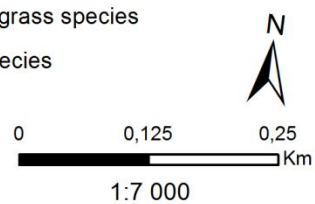
NN classifier



Classes

- | | |
|---|--|
|  Grasslands dominated by <i>Nardus stricta</i> |  Degraded meadows dominated by Dicotyledons |
|  Oligotrophic grasslands |  Degraded meadows dominated by grass species |
|  Mesophile grasslands |  Stands dominated by <i>Vaccinium</i> species |

 Boundary of enclave



Appendix XVIII - Error matrices for the classification of Zadní Rennerovky on the flight line 10_20

Appendix XVIII.A. The error matrix for the SVM classifier. Numbers of classes: 1 – stands dominated by *N.stricta*; 2 – oligotrophic grasslands; 3 – mesophile grasslands; 4 – waterlogged grasslands; 5 – degraded meadows dominated by *Dicotyledons*; 6 – degraded meadows dominated by grass species; 7 – stands dominated by *Vaccinium* species. Abbreviations: OA – overall accuracy, UA – user’s accuracy, PA – producer’s accuracy; CE – commission error; OE – omission error (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	7	Tot(pix)	UA	CE
1	40933 / 79.01%	431 / 15.63%	1871 / 15.81%	14 / 3.02%	722 / 5.69%	4863 / 9.04%	111 / 3.10%	48945	40933 / 48945 83,63%	8012 / 48945 16,37%
2	125 / 0.24%	1582 / 57.38%	71 / 0.60%	0 / 0.00%	130 / 1.02%	323 / 0.60%	0 / 0.00%	2231	1582 / 2231 70,91%	649 / 2231 29,09%
3	412 / 0.80%	105 / 3.81%	3807 / 32.18%	0 / 0.00%	788 / 6.21%	762 / 1.42%	166 / 4.64%	6040	3807 / 6040 63,03%	2233 / 6040 36,97%
4	10 / 0.02%	0 / 0.00%	0 / 0.00%	7 / 1.51%	2 / 0.02%	6 / 0.01%	2 / 0.06%	27	7 / 27 25,93%	20 / 27 74,07%
5	32 / 0.06%	51 / 1.85%	7 / 0.06%	0 / 0.00%	2044 / 16.11%	1060 / 1.97%	1 / 0.03%	3195	2044 / 3195 63,97%	1151 / 3195 36,03%
6	10241 / 19.77%	587 / 21.29%	5975 / 50.50%	441 / 95.25%	8816 / 69.50%	46657 / 86.72%	2935 / 82.10%	75652	46657 / 75652 61,67%	28995 / 75652 38,33%
7	54 / 0.10%	1 / 0.04%	101 / 0.85%	1 / 0.22%	183 / 1.44%	130 / 0.24%	360 / 10.07%	830	360 / 830 43,37%	470 / 830 56,63%
Tot (pix)	51807	2757	11832	463	12685	53801	3575	136920		
PA	40933 / 51807 79,01%	1582 / 2757 57,38%	3807 / 11832 32,18%	7 / 463 1,51%	2044 / 12685 16,11%	46657 / 53801 86,72%	360 / 3575 10,07%		OA = 69.57%	
OE	10874 / 51807 20,99%	1175 / 2757 42,62%	8025 / 11832 67,82%	456 / 463 98,49%	10641 / 12685 83,89%	7144 / 53801 13,28%	3215 / 3575 89,93%		kappa = 0.53	

Appendix XVIII.B. The error matrix for the NN classifier (source: own ENVI analysis).

Pix / %	1	2	3	4	5	6	7	Tot(pix)	UA	CE
1	41231 / 79,59%	621 / 22,52%	2148 / 18,15%	35 / 7,56%	1064 / 8,39%	5736 / 10,66%	278 / 7,78%	51113	41231 / 51113 80,67%	9882 / 51113 19,33%
2	155 / 0,30%	1291 / 46,83%	11 / 0,09%	0 / 0,0%	169 / 1,33%	382 / 0,71%	0 / 0,0%	2008	1291 / 2008 64,29%	717 / 2008 35,71%
3	156 / 0,30%	135 / 4,90%	2646 / 22,36%	0 / 0,0%	387 / 3,05%	567 / 1,05%	270 / 7,55%	4161	2646 / 4161 63,59%	1515 / 4161 36,41%
4	0 / 0,00%	0 / 0,00%	0 / 0,00%	0 / 0,0%	0 / 0,0%	0 / 0,0%	0 / 0,0%	0	0 / 0 0,00%	0 / 0 0,00%
5	90 / 0,17%	68 / 2,47%	193 / 1,63%	0 / 0,0%	711 / 5,61%	471 / 0,88%	2 / 0,06%	1535	711 / 1535 46,32%	824 / 1535 53,68%
6	10142 / 19,58%	642 / 23,29%	6746 / 57,01%	428 / 92,44%	10279 / 81,02%	46589 / 86,60%	2928 / 81,90%	77754	46589 / 77754 59,92%	31165 / 77754 40,08%
7	32 / 0,06%	0 / 0,00%	88 / 0,74%	0 / 0,0%	75 / 0,59%	56 / 0,10%	97 / 2,71%	349	97 / 349 27,79%	259 / 349 72,21%
Tot (pix)	51807	2757	11832	463	12685	53801	3575	136920		
PA	41231 / 51807 79,59%	1291 / 2757 46,83%	2646 / 11832 22,36%	0 / 463 0,00%	711 / 12685 5,61%	46589 / 53801 86,60%	97 / 3575 2,71%		OA = 67.61%	
OE	10576 / 51807 20,41%	1466 / 2757 55,17%	9186 / 11832 77,64%	463 / 463 100,00%	11974 / 12685 94,39%	7212 / 53801 13,40%	3478 / 3575 97,29%		kappa = 0.49	

Appendix XIX – The contents of the attached CD

Folder “text” – contains the text and appendices of the master thesis

Folder “script” – contains the Python scripts utilized in the master thesis