Alas Landscape Modeling by Remote Sensing Image Analysis and Geographic Ontology. Study case of Central Yakutia (Russia)

Sébastien Gadal [1,2][a], Moisei Zakharov [1,2][b], Jūratė Kamičaitytė [3][c], and Yuri Danilov [2][d]

[1] *Aix-Marseille Univ, CNRS, ESPACE UMR 7300, Univ Nice Sophia Antipolis, Avignon Univ,13545 Aix-en-Provence.*

[2] *North-Eastern Federal University, 670000 Yakutsk, Republic of Sakha, Russian Federation.*

[3] *Kaunas University of Technology, Kaunas, Lithuania*

Keywords: Geographic Ontology, Image Analysis, Knowledge Database, Image Processing, Alas Landscape, Remote Sensing

Abstract: Approaches of geographic ontologies can help to overcome the problems of ambiguity and uncertainty of remote sensing data analysis for modeling the landscapes as a multidimensional geographic object of research. Image analysis based on the geographic ontologies allows to recognize the elementary characteristics of the alas landscapes and their complexity. The methodology developed includes three levels of geographic object recognition: (1) the landscape land cover classification using Support Vector Machine (SVM) and Spectral Angle Mapper (SAM) classifiers; (2) the object-based image analysis (OBIA) used for the identification of alas landscape objects according to their morphologic structures using the Decision Tree Learning algorithm; (3) alas landscape's identification and categorization integrating vegetation objects, territorial organizations, and human cognitive knowledge reflected on the geo-linguistic object-oriented database made in Central Yakutia. The result gives an ontology-based alas landscape model as a system of geographic objects (forests, grasslands, arable lands, termokarst lakes, rural areas, farms, repartition of built-up areas, etc.) developed under conditions of permafrost and with a high sensitivity to the climate change and its local variabilities. The proposed approach provides a multidimensional reliable recognition of alas landscape objects by remote sensing images analysis integrating human semantic knowledge model of Central Yakutia in the subarctic Siberia. This model requires to conduct a multitemporal dynamic analysis for the sustainability assessment and land management.

1 INTRODUCTION

The problem of knowledge integration in remote sensing for the studies of landscape characteristics is one of the key gaps; especially for the modeling of the territorial sustainability assessment and ongoing future landscape changes (Konys, 2018). The expert geographic knowledge building based on geographic ontologies is the most suitable approach. It allows the integration of expert-human knowledge for the remote sensing image analysis and modeling of the complexity of the alas landscape system. The approach based on geographic ontology can be considered as a specific domain of knowledge in the field of the artificial intelligence. Geographical ontologies used for the alas landscape modeling includes: (a) biophysical geographic components of landscape (e.g., vegetation, moisture, landform (relief), climate conditions, soil (Pashkevich, 2017)); (b) morphologic forms and geometric objects characteristics (e.g., shape, size, textures, topology, geolocation, depression, inclination, slope), and (c) anthropogenic objects related to the land field reality measured (captured) by remote sensing, the human geographic object's identification (semantic meaning) and perceptions. The human geographic object's identification

[a] https://orcid.org/0000-0002-6472-9955

[b] https://orcid.org/0000-0002-8916-2166

[c] https://orcid.org/0000-0002-8821-5764

[d] https://orcid.org/0000-0003-4806-4938

(semantic meaning) was formalized in Russian and Yakutsk languages in a geolinguistic database integrating toponyms, territorial knowledge of local population, census, land uses, maps, etc. (Zamorshikova et al, 2018). This geo-linguistic database including the bunch of heterogeneous knowledge poses the problem of standardization (harmonization). The application in the alas landscape combining the knowledge model, remote sensing and geographic ontologies is based on the theoretical model of ontology (Sinha and Mark, 2010) and on the methodology developed in remote sensing applied to the subarctic urban environment in Yakutsk (Gadal and Ouerghemmi, 2019).

2 ISSUES OF CLIMATE CHANGES OF THE ALAS LANDSCAPE

The interest to study the alas landscape is related to the impacts of climate change and modeling the landscape processes. The alas landscapes are localized in Taiga boreal forest and permafrost territories of the subarctic region of Siberia, Alaska and Canada. They are formed as a result of thawing of the permafrost layer in the exposed parts of the taiga. It makes the alas one of the key indicators of the permafrost degradation and vulnerability of landscapes in the subarctic region (Fedorov, 2019) and has a major contribution on carbon (CH_4 and CO_2) emission. They have impact on the forms of steppe azonal soils which have a high bio-productivity for the agriculture. Agriculture in the subarctic regions is the main framework of the traditional local lifestyles of Yakutsk and Ekenkii. According to it, the cultural, social and territorial significance of the alas is defined (Desyatkin et al, 2018).

Therefore, the modeling of the alas landscape needs to integrate both the human semantic factors and the geo-physical reality, associating quantitative and semantic parameters defining ontologies of the alas geographic space. The approach for recognizing the geographic ontologies merges manually collected data from the geographic landscape (field observations, population knowledge formalized in geo-linguistic database), maps and remote sensing data: Landsat series, 5 TM, 7 ETM+, 8 OLI and Sentinel 2. Geographic object ontologies are extracted from the satellite images by semi-automatic or automatic approach of image analysis (Clouard et al, 2013) using machine learning algorithms Spectral Angle Mapper (SAM) on the Landsat series images according to the better performance (overall accuracy). Two approaches were applied for the Sentinel 2 images: (a) the Support Vector Machine (SVM) for the recognition of the geographic objects by pixel-based classification (Li et al 2009, Ouerghemmi et al, 2017) showing the land cover landscape of the alas and the environment around, and (b) the OBIA based on the Decision Tree algorithm (DT) (Vopham et al, 2018) and the morphological image filtering based on the contrast and texture information through large-scale smoothing for image segmentation (Maragos, 1989, Sofou et al, 2005) that helped to identify automatically the current different geographic objects defining the actual alas. Results are integrated and merged in GIS. It must be noted that the application of the DT algorithm on the Landsat TM, ETM+, OLI is not relevant because of the too low GSD of 30m.

2.1 Study area

The Central Yakutia lowland is characterized by the widespread distribution of Alas. Throughout the territory there are about 16000 alas between Vilyuy, Lena and Aldan Rivers. The study area is located at 40 km distance to the North from the city of Yakutsk covering the surface of 30000 hectares (Fig. 1). The study area is a boreal forest represented by pine and larch. The soil basis is permafrost with a layered and lenticular texture, where thermokarst cryogenic processes are developed, which cause the formation of numerous alas.

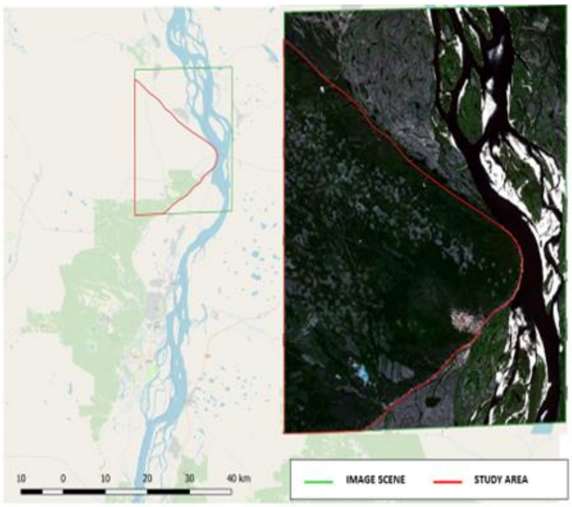

Figure 1. The study area [© Open Street Map]; the extent of the Sentinel 2A RGB scene at 26th of July 2019.

3 IMAGE ANALYSIS OF ALAS LANDSCAPE STRUCTURE

3.1 Methodology

The geographic objects as ontological elements are the basis of the structure and organization of a landscape (Gadal, 2012). In remote sensing they are identified by image processing and image analysis (here by machine learning, landscape object classifications, and OBIA image segmentation) with the knowledge GIS database and the implementation of ontological rules (Fig. 2).

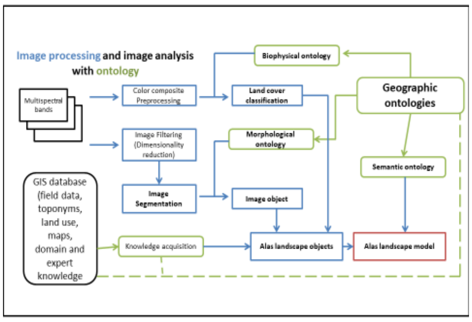

Figure 2. Implementation of image analysis with geographic ontologies for alas landscape recognition.

The typical alas landscapes consist of the following elementary geographic objects: lake, grassland, woodland, built-up and arable lands for agriculture. First, they are identified and analyzed according to the land cover classes produced from the biophysical parameters measured spectrally by Landsat series and Sentinel 2 data.

The second level of the geographical objects recognition by object-oriented classification algorithms of DT is resulted from knowledge acquisition (Clouard et al, 2010) structured in the geo-linguistic database and integrated in GIS, and the OBIA to extract the morphology structures of the alas landscape. The categorization of the alas landscapes objects generated is followed-on by the fusion of land cover maps and alas landscape object classification.

The data analyzed include the expert and the domain knowledge (landscape theories, concepts, indicative and cognitive analysis products) (Gadal and Ouerghemmi, 2019). The ontological approach enables us to obtain numerous representations of all aspects of the alas landscape spatial organization through the accurate acquired knowledge structured in the GIS database. It is used both for training, learning and the validation of results, the spatial analysis and modeling in GIS.

The integrity and the logic coherence of all processes taking place in geographic space is ensured by the integration of semantic geographic object ontologies modelled by the geostatistical analysis of geo-linguistics database (toponyms, territorial knowledge of local population, census, land uses, maps, etc.) with the categorization of alas landscape based on GIS knowledge database.

The obtained alas landscape model (Fig. 3) is a multidimensional model including knowledge and geo-linguistic GIS database adapted for machine learning training and simulation of geospatial processes: rural exodus, permafrost degradation, etc. having impact on the landscape's changes and their consequences (sustainability, re-greening).

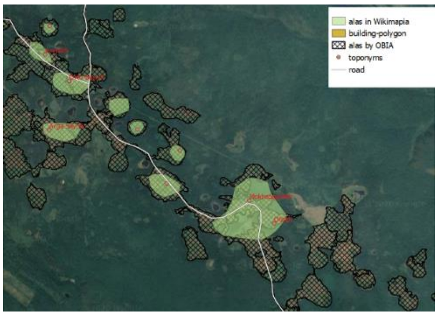

Figure 3. Alas landscape model map with knowledge and geo-linguistic GIS database. Terra-Metrics, DigitalGlobe 2012.

3.2 Image processing

3.2.1 Recognition of objects of landscape land cover change

The recognition of the landscape objects and the change dynamics modeling between 2007 and 2019 is based on the use of the Landsat 5 TM (date: 15.07.2007), Landsat 8 OLI (bands 2 to 7 : VIR-NIR-SWIR) pansharpened at 15m (date: 26.07.2013) and Sentinel 2A (bands 2 to 5 : VIR- NIR at 10m) (date: 24.07.2019) acquired in July.

Landscape land covers produced and mapped by supervised classification are used as training data of the GIS database. Landscape objects are recognized by the spectral signatures related to the geographic objects of the GIS database. Several classifier algorithms were tested: Minimum Distance (MD), Maximum Likehood(ML), the K-Nearest Neighbors (K-NN), Spectral Angle Mapper (SAM) and Support Vector Machine (SVM).

The results show that the optimal machine learning classifiers are obtained with SAM for the Landsat 5 TM and the Landsat 8 OLI data confirming the results presented by Pertopoulos et al. (2010) (Tab. 1).

The most reliable results for the landscape land cover were proceeded with the SVM based on Kernel method for the Sentinel 2A images (Pal and Mather, 2005) (Fig. 4, 5).

Table 1. Results obtained from applied classifiers.

Classification algorithm	O.A. / Kappa		
	2007	2013	2019
MD	72%/0,45	67%/0,44	45%/0.37
ML	43%/0,31	56%/0.39	54/0.36
K-NN	non	non	34%/0.21
SAM	77%/ 0.57	81%/ 0.64	67%/0,59
SVM	non	non	83%/0,65

The accuracy of landscape land cover classifications is provided with 50 randomly generated polygons (3x3 pixels size). The random set of vector data (by 3x3 pixels) constitutes the validation data set. They are used for the accuracy assessment of the classification by confusion matrix and Kappa coefficient.

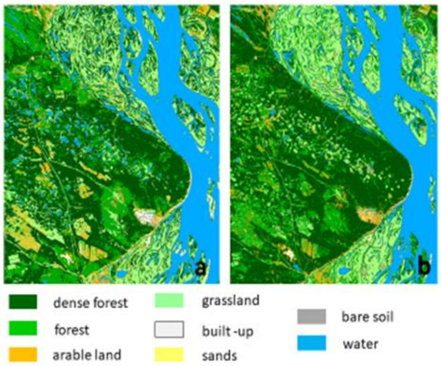

Figure 4. Land cover classification by SAM for Landsat 5 TM in 2007 (a) and Landsat 8 OLI in 2013 (b).

Figure 5. Land cover classification result for Sentinel 2A with SVM classifier.

3.2.2 Alas recognition by object-based image analysis

The recognition of alas is based on two criteria that ontologically characterize geographical objects: morphology and their own geometries. They lend themselves to OBIA procedures. The satellite images used are from the VIR and NIR bands of Sentinel 2A, in particular because of their spatial resolution of 10m that enables to recognize the geometric shapes and the structures of the alas. The alas as a unique geographic object is characterized ontologically by its geometry and by the geographic objects that compose it such as lakes, meadows, swamps, the presence of farms and forests surrounding it. These geographic objects are identified by classification (part 3.2.1.) with the landscape cover maps (fig. 4, fig. 5).

The detection, identification and extraction of alas from their morphological characteristics lends itself well to the OBIA object-oriented machine learning classification algorithms. The OBIA alas recognition consists of four phases: (a) The dimensionality reduction by the Principal Component Analysis (PCA) algorithm, and linear combination of original bands that contains 91-95% of spectral information in first three components (Fig 6a). (b) Smoothing filtering that is applied for the suppression of noise and homogenize statistically the alas regions. This filtering method associates with each pixel of the image the closest local mode in the density distribution of the combined region (Comaniciu and Meer, 1999, 2002). (c) The image segmentation by Large Scale MeanShift non-parametric and iterative clustering method (Michel et al, 2015). The segmentation produces a labeled image with tile-wise processing where neighbor pixels whose range distance is below range radius and optionally spatial distance below spatial radius are merged into the same raster value (Fig. 6c). The average spatial neighborhood radius is 25 meters.

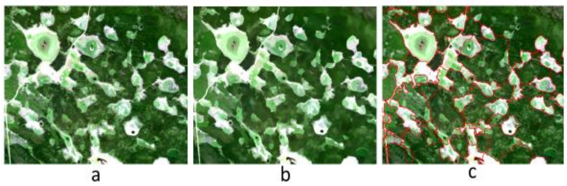

Figure 6. Object recognition from Sentinel 2A (a) RGB of PCA components; (b) MeanShiftSmoophing filtering; (3) Image segmentation of alas.

The most complete extraction of alas is obtained with the following parameters: the spatial radius – 50; the range radius – 25; the Minimum segment size – 10. The main geometrical parameters distinguishing alas and thermokarstic basins concern their shapes: alas has a shape close to the circle, does not have distinct angles, and is not characterized by elongation contrary to thermokarstic shapes. (d) The fourth phase consists in separating the alas from another anthropogenic objects: agricultural parcels. The identification methodology uses a decision tree algorithm. This machine learning algorithm required a dataset that selected 30 polygons generated using the GIS database, more specifically the alas toponyms, topographic maps and Open Street Map data. The classification structure defined by the decision tree is estimated from the training data using a statistical procedure of overall mean and standard deviation for 3 PCA components. The decision rules developed are used to associate each segment of the image with one of the geographic object classes. The O.A. with Sentinel 2A data is =88.4%; a Kappa coefficient is 0,83 for the Decision Tree classifier.

The result is two broad types of geographic objects in which elementary ontological and geographic structures are reflected: alas and arable lands of agriculture (Fig. 7). The information was obtained through the interpretation of remote sensing data by artificial intelligence with the integration of expert knowledge for learning and validation of the obtained results. This GIS-structured information set provides the basis for a comprehensive spatial analysis using semantic ontology to model the landscape dynamics of the boreal forests of alas and permafrost.

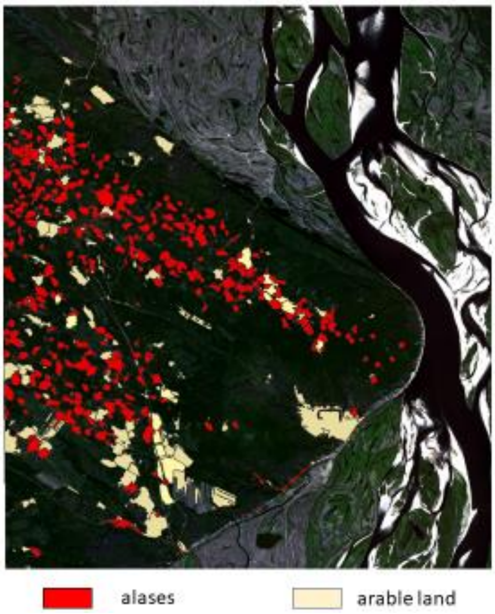

Figure 7. Alas identification by OBIA with Sentinel 2A image.

4 ALAS LANDSCAPE MODELING

In this part, some applications of the image analysis data obtained in the modeling of the alas landscapes are presented. The ecological status of the alas landscapes plays an important socio-economic role, as well as in the dynamic processes of the permafrost. The results of spatial remote sensing image analysis using GIS-integrated geographic ontologies are designed to serve as a basis and environment for this research. Thus, the modeling of geo-linguistic landscape features has been developed in GIS with the creation of an extensive geographically referenced database. It aims to understand better the land use and to measure the complexity of landscape structure.

The first method consists of performing modeling analysis land covering changes within alas. The four categories were identified according to the structure of alas vegetation: (a) Termokarst lake, when the alas basin is completely filled with water; (b) Complex alas, when vegetation changes from steppe to coastal-aquatic; (c) herbaceous alas, when vegetation is relatively uniform throughout all alas territories; (d) pasture alas, with active arable or hay areas. The application of knowledge about the vegetation species of alas allows interpreting land cover classes and to assess the stability and transformation of the alas landscape. In total, the alas include 5 typical land cover classes. A highly moistened alas section belongs to the *grass*, less – often to the *forest* due to close vegetative indicators (chlorophyll intensity). In some cases, the presence of forests inside the alas indicates the presence of a hill, that are indicators of the presence of permafrost heaving tubercles – pingo. The *arable land* is the middle alas section, the most bio-productive section often used as a forage base for cattle-breeding and horse-breeding. The peripheral portion of alas with a lack of moisture forms the azonal steppe vegetation and stands out as a class of *bare soils*. Table 2 shows examples of alas with the percentage structure of the land cover and the corresponding category.

This categorization allows us to consider the stage of spatio-temporal development of alas, water content, and to obtain data on anthropization and economic activity. This model also has a perspective for the development of simulation strategies for the transformation of plains landscapes of permafrost.

Table 2. Examples of alas landscape categorization by land cover structure.

Name	Year	Land cover classes (%)					Alas category
		forest	Arable land	grass	bare soil	water	
Ton-Bas (Тон-Бас)	2007	17	2	34	0	48	termo-karst lake
	2013	18	6	23	17	35	complex alas
	2019	5	10	63	20	2	complex alas
Mokhsogolokh (Мохсоголлох)	2007	53	0	9	0	37	termo-karst lake
	2013	1	0	99	0	0	herba-ceous alas
	2019	5	38	57	0	0	pasture alas
Sier-Bie (Сиэр-Бие)	2007	1	0	81	0	18	herba-ceous alas
	2013	1	45	53	0	1	pasture alas
	2019	1	47	42	0	11	pasture alas

The second method is based on landscape semantic analysis using the geo-linguistic database. The alas landscapes are the main source of livelihood for the local population, which implies the presence of a cultural layer of information of the geographical ontology, within the framework of local knowledge about the geographical space. Semantics provides a transition from the object level of the geographical ontology, where the main concept is geographical objects represented in our case as land-use classes - the alas - and the results of their modeling according to the expert knowledge (alas categories, vegetation species), to the spatial level of the geographical ontology. This territorial knowledge generates geographic models recognition and identification of which constitute a new challenge for artificial intelligence in space and airborne remote sensing.

The spatial organization of the local economy involves the location of villages and farms in the alas. Due to their rarity, they cannot be recognized by image processing from medium spatial resolution images such as Sentinel 2A/B, Landsat 5/6/7/8, and are included in the class of arable (agricultural) land. When their presence was known, we reclassified the pixel sets of the arable (agricultural) land value into the buildings class by comparing them with the semantic layer of settlements and sayylyks (summer houses).

An initial geo-linguistic analysis also allows us to identify the alas used in agriculture, which may have been abandoned recently. Compared to satellite images analysis, the absence of a semantic layer can indicate the age and time of alas formation. The GIS semantic database also allows the dissemination of remote sensing knowledge and facilitates the interpretation of landscape ontology. It provides a semantic description of the results of analysis and processing of remote sensing images.

5 CONCLUSIONS

Geographic ontologies image analysis in artificial intelligence is a necessary part of knowledge that combines multi-level and heterogeneous concepts of an object and a geographic space, and as well geo-linguistics in order to implement a systematic approach to the modeling of alas landscapes. Although, the developed model confronts the technical issues, such as the standardization and harmonization of heterogeneous information flows, it also establishes the rules for solving decision-making problems, for environmental and agricultural monitoring under the conditions of continuously expanding permafrost in subarctic climate.

This research significantly contributes to the development of application of knowledge model to improve remote sensing image processing analysis and landscape modeling for the territorial and geographic studies. The novelty of the research comprises the integration of the geo-linguistic database reflecting the semantic meaning of people's knowledge as part of the geographic ontology of alas landscapes. It allows to analyze the value, the age and the evolution of alas increasing the accuracy of recognition of alas objects.

One of the main challenges is to hybridize geographical knowledge with artificial intelligence used to analyze the landscapes and the cross processes of anthropization and geophysical evolution of permafrost by spatial and airborne remote sensing.

ACKNOWLEDGEMENTS

This research was supported by the French National Research Agency (ANR) through the PUR project (ANR-14-CE22- 0015), the Russian Science Foundation № 15-18-20043 “Landscape Ontology: Semantics, Semiotics and Geographical Modeling”, and the Vernadski Grant of the French Embassy in the Russian Federation.

REFERENCES

Clouard, R., Renouf, A., Revenu, M., 2010. An Ontology-Based Model for Representing Image Processing Objectives. *International Journal of Pattern Recognition and Artificial Intelligence, World Scientific Publishing,* pp.1181-1208.

Comaniciu, D. and Meer, P. 1999. Mean shift analysis and applications. International Conference on Computer Vision, 2, pp. 1197-1203.

Comaniciu, D., and Meer, P. 2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on pattern analysis and machine intelligence, 24(5), pp. 603-619.

Desyatkin, A., Takakai, F. and Nikolaeva, M. 2018. Landscape Microzones within Thermokarst Depressions of Central Yakutia under Present Climatic Conditions. *Geosciences 2018,* 8. MDPI

Fedorov, A., 2019. Permafrost Landscapes: Classification and Mapping. *Geosciences 2019, 9*. MDPI.

Gadal, S., 2012. Geographic Space Ontology, *Locus Object, and Spatial Data Representation Semantic Theory.* Universal Ontology of Geographic Space: Semantic Enrichment for Spatial Data, IGI Global, pp.28-52, 9786-1-4666-032-27-1.

Gadal, S., and Ouerghemmi, W., 2019. Knowledge Models and Image Processing Analysis in Remote Sensing: Examples of Yakutsk (Russia) and Kaunas (Lithuania). *In Proceedings of the 5th International Conference on Geographical Information Systems Theory, Applications and Management,* Heraklion, Crete – Greece. pp. 282-288.

Konys, A., 2018. An Ontology-Based Knowledge Modelling for a Sustainability Assessment Domain. *In Sustainability Journal 10.*

Kordjamshidi, P., Moens, M-F. 2014. Global Machine Learning for Spatial Ontology Population. *Journal of Web Semantics* pp. 1-29

Li, J., Carlson, B.E., and Lacis A.A., 2009. A study on the temporal and spatial variability of absorbing aerosols using total ozone mapping spectrometer and ozone monitoring instrument aerosol index data. *J. Geophys. Res., 114.*

Maragos, P. 1989 A representation theory for morphological image and signal processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11.6* pp. 586–599.

Michel, J., Youssefi, D., & Grizonnet, M. (2015). Stable mean-shift algorithm and its application to the segmentation of arbitrarily large remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 53(2), pp. 952-964.

Ouerghemmi, W., Gadal, S., Mozgeris, G., Jonikavičius, D., and Weber, C., 2017. Urban objects classification by spectral library: feasibility and applications. In 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, pp. 1- 4.

Sinha, G., Mark, D., 2010. Toward A Foundational Ontology of the Landscape. MDPI.

Sofou, A., Evangelopoulos, G., Maragos, P. 2005. Soil Image Segmentation and Texture Analysis: A Computer Vision Approach. *IEEE Geoscience and Remote Sensing Letters, Vol. 2, No. 4, pp. 394-398*

Pal, M. and Mather, P. 2005. Support vector machines for classification in remote sensing. *In International Journal of Remote Sensing Vol.5*. pp. 1007-1011

Pashkevich, M., 2017. Classification and Environmental Impact of Mine Dumps. *In Assessment, Restoration and Reclamation of Mining Influenced Soils.* pp. 1-32

Petropoulos, G., Vadrevu, K., Xanthopoulos, G., Karantounias, G., Scholze, M. 2010. A Comparison of Spectral Angle Mapper and Artificial Neural Network Classifiers Combined with Landsat TM Imagery Analysis for Obtaining Burnt Area Mapping. *Sensors.* pp. 1967-1985.

Puebla-Martınez, M., Perea-Ortega, J., Simon-Cuevas, A., and Romero, F. 2018. Automatic Expansion of Spatial Ontologies for Geographic Information Retrieval. *Communications in Computer and Information Science.* pp. 659-670

VoPham, T., Hart, J., Laden, F. and Chiang, Y. 2018. Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environ Health 17, 40*

Zamorshchikova, L., Khokholova, I., Ikonnikova, A., Samsonova, M. and Lebedeva, V. 2018. Toponymic Landscape of Central Yakutia: Etymological Analysis of Geographical Names. *Journal of Siberian Federal University. Humanities & Social Sciences*.