

Article

Not peer-reviewed version

Forest Communities Spatial Modeling as a Basis for Assessing the Sequestration Potential of Ecosystems (Republic of Tatarstan, Russia)

[Artur Gafurov](#)^{*}, Vadim Prokhorov, Maria Kozhevnikova, [Bulat Usmanov](#)

Posted Date: 1 November 2023

doi: 10.20944/preprints202311.0008.v1

Keywords: unsupervised classification; forest communities; carbon balance; remote sensing, google earth engine



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Forest Communities Spatial Modeling as a Basis for Assessing the Sequestration Potential of Ecosystems (Republic of Tatarstan, Russia)

Artur Gafurov *, Vadim Prokhorov, Maria Kozhevnikova and Bulat Usmanov

Institute of Environmental Sciences, Kazan Federal University, 420097 5 Tovarisheskaya str, Kazan, Russian Federation; vadim.prokhorov@gmail.com (V.P.), mania_kazan@mail.ru (M.K.), busmanof@kpfu.ru (B.U.)

* Correspondence: amgafurov@kpfu.ru

Abstract: For the territory of the Republic of Tatarstan (Russia), a spatial model of forest communities was built as a basis for assessing the sequestration potential of ecosystems. Combination of multi-temporal and multi-spectral satellite imagery from the Landsat 8 and Landsat 9 platform as input data and Google Earth Engine cloud platform were used. The set of 292 vegetation indices and metrics computed from the pre-processed imagery, were combined into dataset. The Weka X-Means clustering algorithm was trained and applied to study area. The unsupervised classification was carried out by vegetation classes in the Braun-Blanquet system. The results of unsupervised classification were verified using data from more than 17,000 relevés with geographic references from the Flora database. For automatic classification, the EuroVeg Checklist expert system in the JUICE 7.1 package was used. The proposed methodology for obtaining initial data and unsupervised classification, supported by an automated expert system, made it possible to obtain a picture of the vegetation distribution in the study area with sufficient accuracy, and in the future, it will be used to assess the sequestration potential of the ecosystems of the region under study.

Keywords: unsupervised classification; forest communities; carbon balance; remote sensing; google earth engine

1. Introduction

Terrestrial and water ecosystems play a significant role in carbon sequestration through the creation and retention of organic carbon mass within plant biomass and soil. To date, there exists a plethora of publications dedicated to various aspects of this issue. A comprehensive review of numerous methods and practical cases is provided in the work by Kurbatova [1]. A common premise across these studies is the identification of desired parameters for individual objects (for example, carbon content in the wood of a specific tree species), which are then extrapolated to groups of objects (such as all trees of this species in similar ecological conditions) and multiplied by the number of trees to estimate the overall carbon stock in a forest stand. In this case, the commonality among all objects in the group is the plant species concept, which doesn't pose difficulties, since the taxonomy of higher vascular plants is well-established for the most part.

When referring to a volumetric area object, such as a soil layer, a series of samples, where carbon content is determined in the laboratory, is extrapolated to the entire area occupied by this type of soil, assuming equal thickness (for instance) of the humus horizon. Here, two problematic aspects emerge: object classification (in our case, soil classification), and the determination of their area distribution (direct mapping or mapping through decryption) [2,3].

When addressing the carbon exchange between the entire ecosystem as a whole and the atmosphere, similar to the previous example, it is necessary to classify ecosystems in one way or another, and then determine the distribution of each class for area calculation.

Thus, one of the "cornerstones" in creating carbon balance models is the classification of objects for extrapolation of field observations. Vegetation changes over time and space as a result of ecological processes acting on plant populations and communities at different temporal and spatial scales, making it one of the integral indicators of the ecosystem as a whole [4]. This is because the vegetation component of ecosystems is formed depending on both abiotic factors (moisture, light, soil fertility) and anthropogenic influences. Hence, we propose utilizing the spatial model of vegetation cover as a basis for evaluating the sequestration potential of ecosystems.

Recently, renewed interest in vegetation classification has emerged worldwide, and efforts have been undertaken both nationally and internationally to develop new classification systems using standardized procedures [5,6]. Moreover, there's a growing interest in harmonizing approaches and standardizing the informational content of classifications [7]. This interest is driven by the practical necessity of using vegetation typologies and the increasing recognition of their scientific bases [8]. The classification problem is closely tied to assigning a geobotanical description of a particular community to a syntaxonomic unit of an existing vegetation classification. Experience exists in solving this task using expert systems [9].

Very recently, ideas have emerged regarding the amalgamation of vegetation classification procedures and the solutions provided by expert systems, employing the apparatus of fuzzy sets [10]. The proposed algorithms align with the notions of continuous vegetation cover and the individual behavior of plant species concerning environmental factors.

Therefore, all these research directions aim to minimize the subjectivity of researchers at all stages of creating vegetation cover maps from the development of classification to assigning a specific geographical point to one type of ecosystem or another.

The analysis of terrestrial ecosystems, particularly the classification and mapping of vegetation, is a pivotal task for managing natural resources [11]. Vegetation serves as the foundation for all living beings, playing an indispensable role in various ecological and environmental processes. In recent years, remote sensing technologies have substantially evolved, offering a robust tool for accurate and efficient vegetation mapping. A critical aspect of leveraging remote sensing data for vegetation mapping is the choice of classification approach, which largely dictates the accuracy and reliability of the resulting maps. Among the various classification approaches, unsupervised classification has emerged as a vital technique in minimizing researcher subjectivity at all stages of vegetation mapping, from developing a classification scheme to assigning a particular geographic point to a specific ecosystem type [12–16]. Unlike supervised classification, which relies on prior knowledge and extensive field samples, unsupervised classification operates without any a priori knowledge, solely depending on spectrally pixel-based statistics [13]. This inherent characteristic of unsupervised classification significantly reduces the potential bias that may arise from researchers' prior knowledge or assumptions, thereby enhancing the objectivity of the classification process. Unsupervised classification encompasses a variety of clustering algorithms, including rule-based, neural network, and SVM-based approaches, each with its unique advantages and applications in different scenarios. The utilization of unsupervised classification is not confined to preparing for supervised classifiers but extends to visualizing and monitoring similar areas in a scene, capturing a wide range of land use signatures over areas with extensive and unknown ranges of vegetation types. This broad capture capability underscores the versatility and efficacy of unsupervised classification in handling diverse ecosystems and varied vegetation types. Furthermore, the cost-effectiveness of unsupervised classification, in terms of both time and financial resources, is another significant advantage, especially when compared to supervised classification that necessitates a large number of field samples. The reduced need for extensive field sampling is particularly beneficial in scenarios such as agricultural production management, where timely access to spatial distribution information of crops is of great importance [17,18].

In light of the above, optimized unsupervised classification techniques have been proposed to further enhance the accuracy and efficiency of vegetation mapping. For instance, a two-step approach combining Jenks classification and Agglomerative Hierarchical Clustering (AHC) has been implemented to derive vegetation cover types from SAVI data [19]. Such advancements in

unsupervised classification techniques are indicative of the ongoing efforts to minimize researcher subjectivity, thereby contributing to more accurate and reliable vegetation mapping.

The burgeoning interest and continuous advancements in unsupervised classification underscore its importance in the realm of vegetation mapping [17]. By minimizing researcher subjectivity, unsupervised classification paves the way for more objective, accurate, and reliable vegetation maps, which are indispensable for a myriad of ecological, environmental, and agricultural applications. The exploration and adoption of unsupervised classification techniques, coupled with the integration of other emerging technologies, hold promise for further advancing the field of vegetation mapping, fostering a deeper understanding of terrestrial ecosystems, and facilitating more informed decision-making in natural resource management [20].

The main goal of this study is to create a spatial model of forest vegetation in the classes of the Braun-Blanquet system. This classification approach is based on information about the composition and abundance of plant species in a community. Each species is characterized by its own functional characteristics, which ultimately can give a knowledge about a particular community productivity, which in turn can be used to judge the sequestration capacity of ecosystems.

2. Materials and Methods

2.1. Study Area

The study area is located within the Republic of Tatarstan (Russian Federation) situated in the eastern part of European Russia (Figure 1). The territory is divided by river valleys of the Volga, Kama and Vyatka into natural areas [21] including landscapes related to the Eastern European boreal subtaiga and Eastern semi-humid broadleaved forests, the boundary between them being determined by the isoline of hydrothermal coefficient of 1 [22]. In zonal terms, the northern part of the territory belongs to the ecoregion of Sarmatian mixed forests, the southern part to the ecoregion of Eastern European forest-steppes [23].

The area includes various landscape types and vegetation units ranging from the formation of the southern taiga to the fragments of typical steppe [24]. In the river valleys on sandy or sandy-clayey substrate various types of pine forests are formed. Meadow formations have developed as a result of forest clearing and mowing. Steppe has remained mostly on the steep rocky slopes that are not suitable for agricultural development. The predominant types of wetlands are fens.

The flora and vegetation of the Republic of Tatarstan have been studied mainly by researchers from Kazan University. This university has a long tradition of floristic and vegetation studies since the 19th century in connection with S. Korshinsky, A. Gordyagin and M. Markov. Much phytosociological data has been collected during more than 120 years by various authors [25].

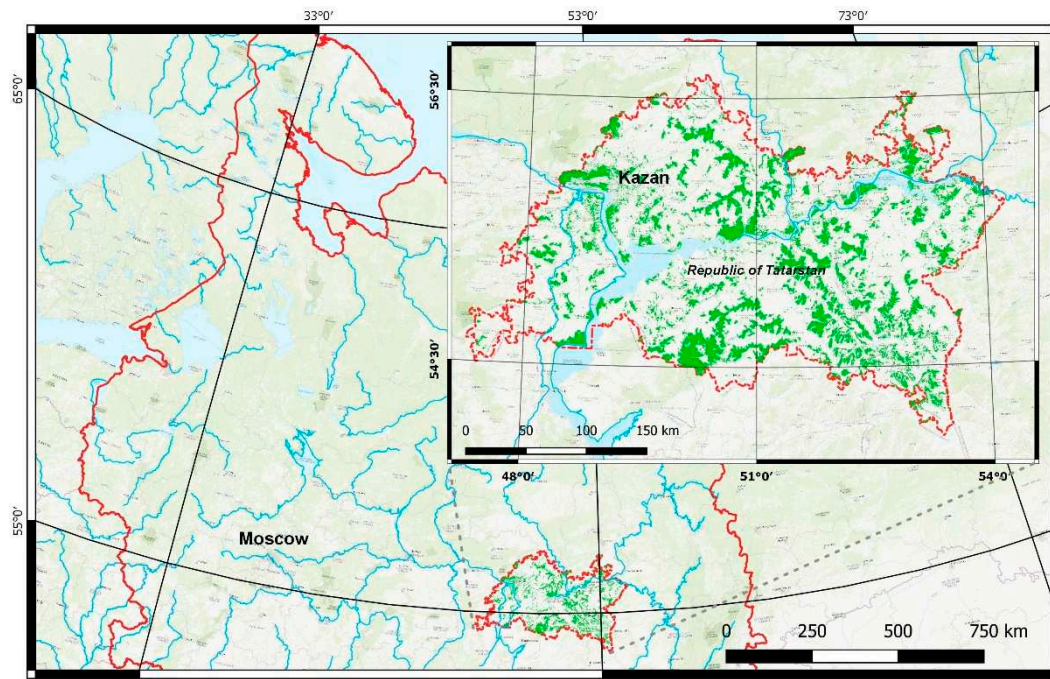


Figure 1. Study area. Forests distribution on the territory of the Republic of Tatarstan.

2.2. Data Acquisition

The data acquisition phase of this study primarily revolved around the utilization of multi-temporal satellite imagery from the Landsat 8 and Landsat 9 platforms. These satellites are part of the established Landsat program, which has been a critical resource for earth observation, aiding in various environmental monitoring [26] and resource management projects [27].

Landsat 8, launched by NASA in 2013, is equipped with two sensors: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). The OLI captures imagery across nine spectral bands, covering visible, near-infrared, and shortwave infrared wavelengths, with a spatial resolution of 30 meters for most bands and 15 meters for the panchromatic band. Conversely, the TIRS captures thermal infrared imagery with a resolution of 100 meters, which is resampled to 30 meters for product delivery.

Landsat 9, akin to Landsat 8, continues the mission of providing high-quality imagery for earth observation. With similar sensors, Landsat 9 ensures a consistent data stream, which is indispensable for long-term environmental monitoring [28].

Several advantages underpin the choice of Landsat imagery for this study. The temporal resolution, with a revisit time of 16 days for both Landsat 8 and Landsat 9, provides an 8-day revisit at the equator when combined. This frequency allows for the capture of temporal dynamics within the forest communities over the study period from 2018 to 2022. Additionally, the 30-meter spatial resolution of Landsat imagery offers a balanced trade-off between detail and coverage, enabling the analysis of forest structure and composition at a landscape scale. Furthermore, the wide range of spectral bands available facilitates the computation of various vegetation indices crucial for assessing different aspects of vegetation health and structure.

In the image processing phase, the raw Landsat imagery underwent a series of pre-processing steps to correct atmospheric and geometric distortions. This was achieved through the *preprocess* method in the *eemont* [29] extension of Google Earth Engine (GEE) [30], which provides functionalities for atmospheric correction and cloud masking, thereby improving the quality and consistency of the imagery. Temporal filtering was also performed to include data from April to November for the years 2018 to 2022, to account for seasonal variations in vegetation cover and exclude winter months where snow cover could interfere with the analysis. Subsequently, a comprehensive suite of vegetation indices was derived from the pre-processed imagery using the *spectralIndices* method in *eemont*. These

indices serve as proxies for various vegetation attributes and are critical for the unsupervised classification task. The Landsat 8 and Landsat 9 imagery were then merged to form a continuous time-series dataset. This dataset was utilized to compute statistical composites like median, mean, maximum, and standard deviation for each of the vegetation indices across the temporal dimension. These composites encapsulate the central tendency and variability of vegetation conditions over the study period, providing a robust dataset for the unsupervised classification.

2.3. Google Earth Engine Platform

The GEE platform played a quintessential role in the data processing and analysis stages of this study. GEE is a cloud-based platform designed for the planetary-scale environmental data analysis. It provides a robust infrastructure necessary for processing large datasets, which is essential given the volume of multi-temporal satellite imagery utilized in this study. GEE houses an extensive archive of satellite imagery and geospatial datasets, which is continuously updated, thereby providing a rich repository of data for environmental studies. For this investigation, the Landsat 8 and Landsat 9 imagery were directly accessed and processed within the GEE environment, which significantly streamlined the data handling process.

One of the significant advantages of using GEE [31] is its ability to handle and process large volumes of data in a parallelized and efficient manner. The platform employs a server-side processing paradigm, which means that the data does not need to be downloaded to local machines for analysis, thus circumventing the traditional bottlenecks associated with large-scale data processing. This is particularly advantageous for studies like ours, where the analysis spans a large geographical area over an extended period.

The GEE platform supports a variety of data processing and analysis tasks essential for this study. These include image collection filtering, image reduction, computation of vegetation indices, and unsupervised classification. The platform also provides a range of pre-built algorithms and machine learning models, alongside the ability to implement custom algorithms, which was pivotal for the unsupervised classification task undertaken.

Furthermore, GEE provides a user-friendly interface and a powerful scripting environment through its Code Editor, which supports both JavaScript and Python. For this study, the Python programming language was employed for scripting, facilitated by the *ee* and *eemont* libraries. These libraries provided the necessary tools and extended functionalities for image processing, feature extraction, and statistical analysis, all of which were crucial for the successful execution of the unsupervised classification of forest communities.

Moreover, the GEE platform also enables easy sharing and exporting of results, which is crucial for collaboration and further analysis. The results from the unsupervised classification were exported to Google Drive, which allowed for a seamless transition to the post-classification analysis stage of the study.

2.4. Vegetation Indices Computation

The computation of vegetation indices (VIs) is pivotal in harnessing the spectral information embodied in the satellite imagery. VIs serve as proxies for various biophysical parameters and are instrumental in delineating vegetation characteristics. In this study, a gamut of VIs was derived from the pre-processed Landsat 8 and Landsat 9 imagery, leveraging the *spectralIndices* method within the *eemont* library in Google Earth Engine.

In our comprehensive analysis of vegetation indices, we adopted a meticulous approach to categorize and scrutinize various indices based on their underlying characteristics and applications. The vegetation indices were classified into distinct categories, each representing a specific aspect of vegetation characteristics captured via remote sensing (Table 1).

Table 1. Characteristics and applications of vegetation indices

Index name	Description
REFLECTANCE-BASED INDICES	
DVI (Difference Vegetation Index) [32]	Captures the difference in reflectance between the near-infrared (NIR) and red bands.
DVIplus [33]	An extension of DVI, considering additional spectral bands or factors.
GNDVI (Green Normalized Difference Vegetation Index) [34]	Utilizes the green and NIR bands to emphasize chlorophyll content.
GRNDVI (Green-Red Normalized Difference Vegetation Index) [35]	Incorporates green and red bands for enhanced vegetation monitoring.
MNDVI (Modified NDVI) [36]	An adjusted form of NDVI to account for atmospheric and canopy background effects.
NDVI (Normalized Difference Vegetation Index) [37]	One of the most widely used indices, highlighting areas of active vegetation.
RDVI (Renormalized Difference Vegetation Index) [32]	Aims to minimize soil background influences while emphasizing vegetation.
TDVI (Transformed Difference Vegetation Index) [38]	A modified version of NDVI to enhance sensitivity to vegetation changes.
WDRVI (Wide Dynamic Range Vegetation Index) [39]	Offers a wider dynamic range than NDVI, improving sensitivity.
WDVI (Weighted Difference Vegetation Index) [40]	Focuses on minimizing soil background effects by applying a specific weight.
WATER CONTENT INDICES	
DSWI1 (Drought Stress Water Index 1) [41]	Highlights areas undergoing drought stress, indicating lower water content.
GSAVI (Green Soil Adjusted Vegetation Index) [42]	A soil-adjusted index using the green band to minimize soil background influences.
NIRv (Near-Infrared Reflectance of Vegetation) [43]	Captures the NIR reflectance associated with vegetation’s structural characteristics.
NIRvH2 [44]	A variation of NIRv, considering additional factors for enhanced accuracy.
NDII (Normalized Difference Infrared Index) [45]	Helps in assessing vegetation water content.
NDMI (Normalized Difference Moisture Index) [46]	Another index for evaluating water content in vegetation.
NMDI (Normalized Multi-band Drought Index) [47]	Focuses on monitoring drought conditions across various bands.
MSI (Moisture Stress Index) [48]	Highlights moisture stress in vegetation, crucial for drought monitoring.
SOIL-ADJUSTED INDICES	
GDVI (Green Difference Vegetation Index) [49]	Minimizes soil background influences using the green band.
GOSAVI (Green Optimized Soil Adjusted Vegetation Index) [42]	A soil-adjusted index designed for optimal vegetation monitoring.
GRVI (Green Red Vegetation Index) [42]	Utilizes green and red bands for vegetation analysis, adjusting for soil effects.
MSAVI (Modified Soil Adjusted Vegetation Index) [50]	A soil-adjusted index with modifications for better accuracy.
OSAVI (Optimized Soil Adjusted Vegetation Index) [51]	Optimizes the soil adjustment to improve vegetation monitoring in low cover areas.

SARVI (Soil Adjusted and Atmospherically Resistant Vegetation Index) [52]	A soil-adjusted and atmospherically resistant index.
SAVI (Soil Adjusted Vegetation Index) [53]	Adjusts for the influence of soil brightness when vegetation cover is low.
TSAVI (Transformed Soil Adjusted Vegetation Index) [54]	A transformed index for better vegetation representation in diverse conditions.
CHLOROPHYLL INDICES	
GARI (Green Atmospherically Resistant Vegetation Index) [34]	Designed to monitor chlorophyll content while minimizing atmospheric effects.
GCC (Green Chlorophyll Content) [55]	Directly related to chlorophyll content, crucial for assessing vegetation health.
MGRVI (Modified Green Red Vegetation Index) [56]	A modified index to enhance sensitivity to chlorophyll content.
MCARI2 (Modified Chlorophyll Absorption in Reflectance Index 2) [57]	Targets chlorophyll absorption features for accurate monitoring.
MSR (Modified Simple Ratio) [58]	A modified vegetation index to improve sensitivity to chlorophyll content.
MTVI2 (Modified Triangular Vegetation Index 2) [57]	Focuses on enhancing the representation of chlorophyll content.
OCVI (Optimized Chlorophyll Vegetation Index) [59]	Optimizes chlorophyll representation in vegetation monitoring.
VIG (Vegetation Index Green) [60]	Utilizes the green band for chlorophyll monitoring, essential for assessing plant health.
STRUCTURAL INDICES	
CVI (Chlorophyll Vegetation Index) [61]	Represents vegetation structure and chlorophyll content.
DSI (Difference Structure Index) [62]	Highlights structural variations in vegetation.
SR (Simple Ratio) [63]	A basic ratio of NIR to red reflectance, indicating vegetation structure.
NON-LINEAR INDICES	
EVI2 (Enhanced Vegetation Index 2) [64]	An improved version of NDVI, incorporating non-linear enhancements.
GBNDVI (Green Blue Normalized Difference Vegetation Index) [35]	A non-linear index utilizing green and blue bands.
GLI (Green Leaf Index) [65]	Represents vegetation greenness in a non-linear manner.
GEMI (Global Environment Monitoring Index) [66]	A global index for vegetation monitoring with non-linear properties.
RI (RapidEye Vegetation Index) [67]	A specific index for RapidEye satellite data.
SI (Shadow Index) [68]	Represents vegetation shape characteristics.
SEVI (Soil and Atmospherically Resistant Vegetation Index) [69]	Minimizes soil and atmospheric effects using a non-linear approach.
VARI (Visible Atmospherically Resistant Index) [60]	Visible spectrum for vegetation monitoring, applying non-linear corrections.
OTHER INDICES	
BWDRVI (Broadband Width Difference Vegetation Index) [70]	A unique index capturing broadband width differences.
CIG (Canopy Index Green) [71]	Represents canopy structure using the green band.
FCVI (Floating Canopy Vegetation Index) [72]	A special index for floating canopy vegetation.

IAVI (Inverted Attributed Vegetation Index) [73]	An inverted index for enhanced vegetation attribute representation.
IKAW (Kawashima Vegetation Index) [74]	A specific vegetation index developed by Kawashima.
IPVI (Infrared Percentage Vegetation Index) [75]	Utilizes infrared reflectance for vegetation percentage estimation.
MRBVI (Modified Ratio Vegetation Index) [76]	A modified ratio index for improved vegetation monitoring.
MNLI (Modified Non-Linear Vegetation Index) [77]	Incorporates non-linear adjustments for enhanced vegetation representation.
NDDI (Normalized Difference Drought Index) [78]	Focuses on drought monitoring and water stress assessment.
NDGI (Normalized Difference Greenness Index) [33]	Highlights vegetation greenness.
NDPI (Normalized Difference Phenology Index) [79]	Utilized for monitoring vegetation phenology.
NDYI (Normalized Difference Yellowness Index) [80]	Highlights vegetation yellowness, crucial for certain phenological stages.
NGRDI (Normalized Green Red Difference Index) [81]	Utilizes green and red bands for vegetation monitoring.
NRFIg (Normalized Red/Far-Red Index Green) [82]	Focuses on the red to far-red ratio using the green band.
NRFIr (Normalized Red/Far-Red Index Red) [82]	Utilizes the red to far-red ratio in the red band.
NormG (Normalized Green) [42]	Represents normalized green reflectance.
NormNIR (Normalized NIR) [42]	Represents normalized near-infrared reflectance.
NormR (Normalized Red) [42]	Represents normalized red reflectance.
RCC (Red Chromatic Coordinate) [55]	Directly related to chlorophyll content using the red band.
RGBVI (Red Green Blue Vegetation Index) [56]	Utilizes the RGB bands for vegetation analysis.
RGRI (Red Green Ratio Index) [83]	A ratio index using red and green bands.
TGI (Triangular Greenness Index) [84]	Represents vegetation greenness using a triangular approach.
TriVI (Triangular Vegetation Index) [85]	A triangular index for enhanced vegetation representation.

The array of VIs computed in this study furnishes a robust dataset encapsulating a myriad of vegetation attributes. These indices, each with its unique domain of application, collectively contribute to a comprehensive feature set forming the basis for the unsupervised classification of forest communities within the study area. Through the judicious computation and utilization of these VIs, a rich tapestry of information regarding vegetation dynamics over time and across the study area was woven, paving the avenue for an in-depth analysis of forest communities through unsupervised classification.

2.5. Unsupervised Classification

The mapping of forest communities in an unsupervised manner was a focal point in this analysis, aimed at uncovering the intrinsic structure and spatial distribution of vegetation types within the study area. The technique employed was rooted in the Weka X-Means clustering algorithm [69] facilitated by the GEE platform, extending the functionality of the traditional K-means algorithm by adaptively ascertaining the optimal number of clusters within a designated range.

2.5.1. Data Pre-Processing

A cornerstone to the ensuing classification was the meticulous pre-processing of data. The ensemble of vegetation indices computed from prior steps was amalgamated to form a cohesive dataset. Subsequently, statistical reducers were employed to distill the temporal essence of the data by generating median, mean, maximum, and standard deviation composites for each index. This was pivotal in encapsulating the temporal dynamics of vegetation attributes over the study timeframe. To refine the dataset further, a mask was applied utilizing a dataset from the Global Land Analysis and Discovery (GLAD) laboratory at the University of Maryland [86], effectively sieving out non-forested areas and thus honing the focus on forested regions.

2.5.2. Training Dataset Preparation

The basis of the clustering process was the preparation of a robust training dataset. A total of 20,000 pixels were randomly sampled across the study area, each pixel representing a data point with a suite of features encapsulated in the vegetation indices and statistical composites. This dataset, rendered at a spatial resolution of 30 meters, served as the crucible for training the clusterer, providing a diverse yet representative glimpse into the spectral and temporal heterogeneity of the vegetation within the study area.

2.5.3. Clustering Process

With a well-curated training dataset at hand, the stage was set for the clustering process. The Weka X-Means clustering algorithm was trained, iterating through a specified range of clusters, specifically from 36 to 50, to dynamically ascertain the optimal number of clusters that best encapsulated the inherent structure of the data. Once trained, the clustering model was applied in a stratified manner to each grid within the study area, engendering a separate classified image for each grid. This stratified approach ensured a localized analysis, catering to the unique vegetative composition and spatial heterogeneity inherent within each grid.

2.5.4. Mosaic Creation and Export

Post-classification, the separate classified images for each grid were melded into a single coherent image through a mosaicking process. This process, facilitated by the mosaic method in GEE, seamlessly stitched together the classified images based on the region boundaries, yielding a comprehensive classified image of the entire study area.

2.6. Field Data Collection and Processing

The primary source of data for this study validation and verification was procured from the extensive "Flora" database [25], which houses a vast collection of over 17,000 relevés. A selective extraction process executed to acquire relevés documented post the year 2000, which also contained geographical references, ensuring relevance and contemporaneity of the data. For maintaining a standardized taxonomy across the dataset, the nomenclature of plant species was meticulously aligned with the globally recognized "The Plant List" database [70]. A deliberate exclusion of bryophyte species from the selection carried out to maintain a focused approach towards vascular plant communities. The classification of the relevés executed at the level of vegetation classes as delineated in the Braun-Blanquet system [87,88]. Automation played a crucial role in the classification process. The expert systems EuroVeg Checklist [9] employed within the JUICE 7.1 software package [89] to facilitate a precise and efficient automated classification. The integration of this advanced expert system streamlined the classification process, ensuring an objective and methodical analysis of the vegetation data.

This rigorous methodology in data extraction, nomenclature standardization, and classification, bolstered by automated expert system, constituted a robust framework for field data collection and processing. This framework was instrumental in ensuring the accuracy and reliability of the classification, which in turn significantly contributed to a comprehensive understanding of vegetation and habitat distributions, aligning with the broader objectives of this study.

3. Results and discussion

The conducted cluster analysis facilitated the delineation of 44 unique statistically distinguishable classes, which formed the foundation of the raster model of forest communities in the Republic of Tatarstan (Figure 2). In the derived model, pixel values are present, some of which exhibit minimal occurrence, introducing a level of 'noise' into the model, while others are frequently occurring values (Figure 3). Such 'noisy' values either need to be eliminated or associated with another class. To address this issue, an approach based on "tail cutting" was adopted – at the 1 and 99 percentiles, which enabled the filtering out of all minimally occurring classes (values 3, 10, 34, 40).

To correlate the classes of the raster model with the existing Braun-Blanquet classification system, two approaches were employed – automated and expert. For the automated search for correspondences, an approach based on the analysis of frequency tables of values from the raster model and corresponding coordinates of geobotanical descriptions was utilized.

The implementation of automated identification, to determine which values of the raster model correspond to a specific type according to the classification of plant communities, was carried out in several stages. Initially, classes of field data were identified, where a vast variety of raster values were observed, and classes where they were more homogeneous (Table 2, Figure 4). In the subsequent stage, grouping and aggregation were performed. The grouping of raster values was conducted based on the community type according to the Braun-Blanquet classification to unveil common patterns (Figure 2) and explore the possibility that certain raster values are more prevalent in specific classes of field data. Following this, statistical analysis was employed to uncover significant associations between raster values and classes of field data. This allowed for the recognition of several patterns that may indicate a correspondence between groups of raster values and specific classes of field data. The results of this analysis can be reviewed in the Table S1 of the Supplementary Materials.

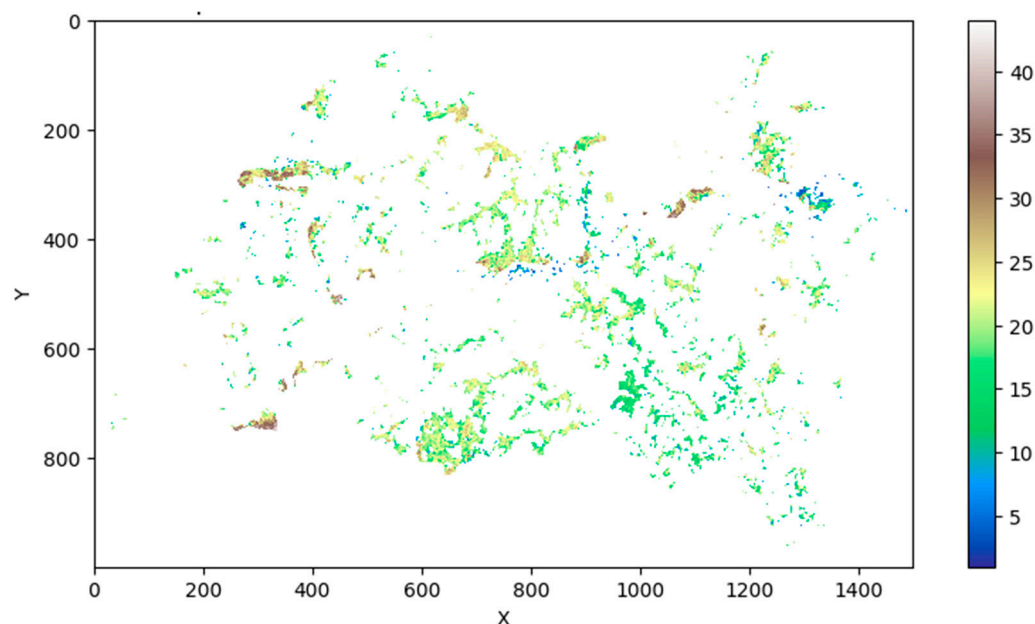


Figure 2. Spatial distribution of forest communities model value.

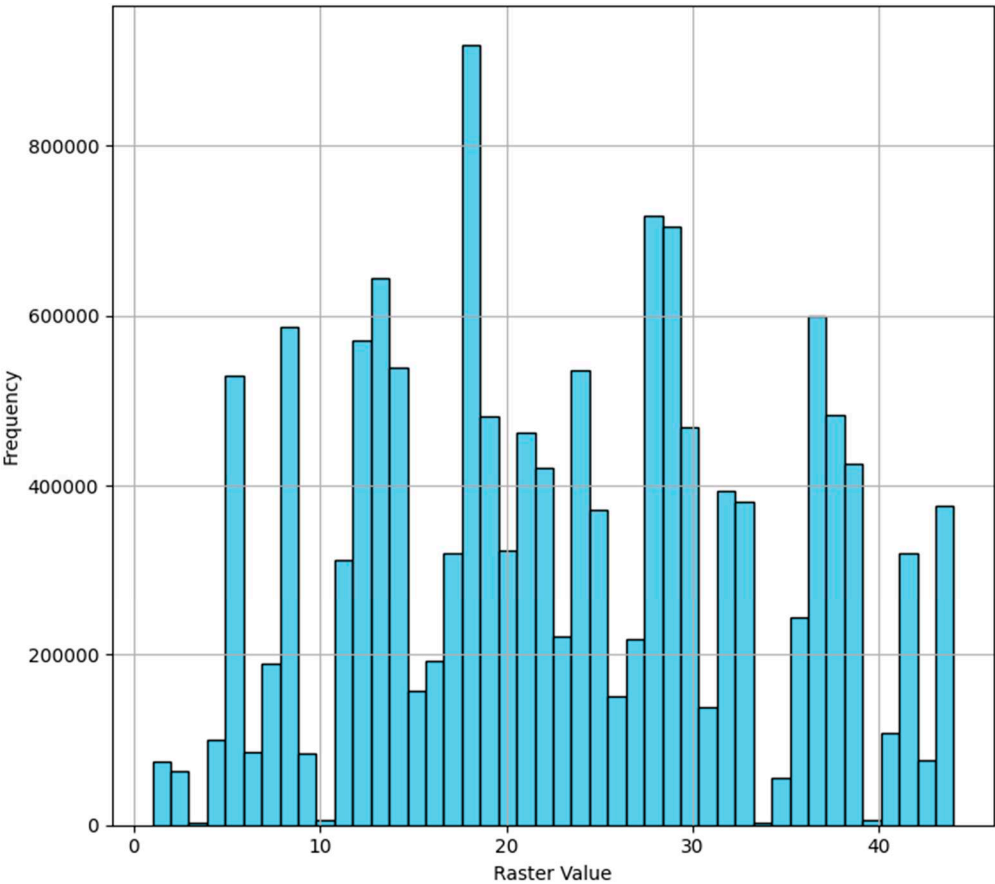


Figure 3. Histogram of raster values of forest communities’ model.

Table 2. Characteristics and applications of vegetation indices.

FC BB*	Main statistics					
	Mean	Median	Std	Min	Max	Count
ALN	28.90	37.00	14.49	5	44	39
ART	22.23	23.00	14.13	1	44	126
ASA	14.67	16.00	9.07	5	23	3
BRA	27.91	39.00	14.66	1	44	464
DIG	12.40	7.00	13.89	1	33	5
EPI	29.47	32.00	11.86	5	44	15
FAG	28.48	33.00	13.22	1	44	1637
FES	23.75	23.00	12.13	1	42	61
FRA	30.00	30.00	NaN	30	30	1
LEM	14.50	14.50	12.02	6	23	2
LOI	39.00	39.00	NaN	39	39	1
LON	12.75	6.50	12.84	6	32	4
MOL	27.23	32.00	12.81	1	44	152
NAR	39.00	39.00	0.00	39	39	3
OXY	29.90	35.50	14.22	2	39	10
PAR	24.67	25.00	17.50	7	42	3
PHR	25.52	30.50	14.83	2	44	64
PIC	32.05	39.00	12.86	2	44	178
POP	30.18	33.00	11.30	5	44	121
POT	24.00	24.00	25.46	6	42	2

PUB	27.00	27.00	NaN	27	27	1
QUE	26.44	39.00	15.30	4	39	25
RHA	25.00	25.00	NaN	25	25	1
SCH	32.00	32.00	NaN	32	32	1
SIS	19.00	28.00	15.59	1	28	3

* Forest communities' class by Braun-Blanquet.

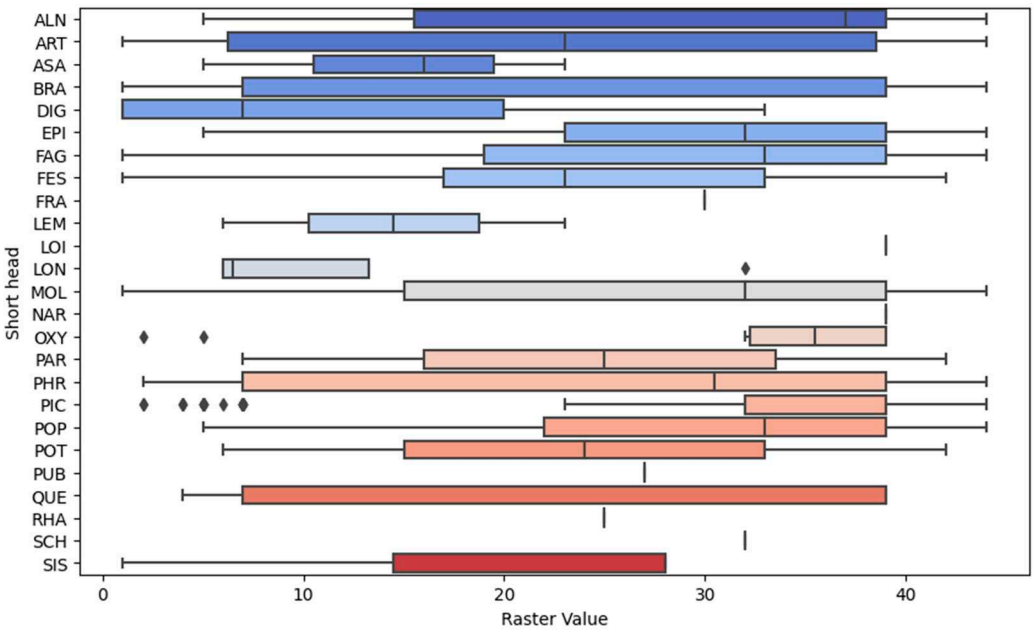


Figure 4. Aggregated raster values across field data classes.

The meticulous procedure of aggregating the raster model values has been instrumental in unequivocally identifying the principal community types, which are denoted as FAG (Carpino-Fagetea sylvaticae), BRA (Brachypodio pinnati-Betuletea pendulae), PIC (Vaccinio-Piceetea), ALN (Alnetea glutinosae), POP (Alno glutinosae-Populetea albae), and PUB (Quercetea pubescentis). It's pertinent to note that the class QUE (Quercetea robori-petraeae) entirely aligns with the FAG group based on the raster model values. The remaining classes either showcase low statistical significance or have been included erroneously due to inaccuracies in the spatial referencing of the geobotanical descriptions. However, these raster model values were not discarded; instead, they were expertly allocated.

Post aggregation, a Z-score assessment was meticulously conducted to measure the deviation of a particular observation from the mean value in a distribution. This is calculated as the ratio of the difference between the observation and the mean value to the standard deviation. The Z-score is pivotal for identifying anomalous values, discerning data percentiles, and determining the position of an observation within the mean value's confidence interval. Additionally, it facilitates data normalization, thus simplifying the analysis and comparison of data. The application of this assessment has enabled a clear correlation of at least one (main) raster model class with a forest community as per the Braun-Blanquet classification. In certain instances, where a high degree of correspondence between multiple raster model classes and one type according to Braun-Blanquet was observed, both raster values were associated with the corresponding type. The remaining values were expertly distributed based on the geobotanical descriptions of areas, aiming to align them with one of the 6 types of forest communities as per the Braun-Blanquet classification found within the study area. In cases of derivative and deadwood forests, the raster model values were categorized under the class OTHERS.

The resultant distribution of pixels across classes of forest communities presented in Table 3. A comprehensive cartographic representation of the raster model of forest communities within the Republic of Tatarstan has been illustrated in Figure 5. This representation is pivotal as it offers a

spatial visualization of the forest community distribution, encapsulating the diverse forest communities within the specified region. It provides a visual aid that significantly contributes to a deeper understanding and interpretation of the forest community distributions and their classifications, aligning with the broader objectives of this study.

Table 3. Classified forest communities' raster model statistics.

Class	# of pixels	Area (km ²)	% of Forested area
ALN	2082534	1874	16
BRA	440568	397	3
FAG	6370003	5733	48
PIC	564690	508	4
POP	876324	789	7
PUB	1320184	1188	10
OTHERS	1742692	1568	13

For the validation of the obtained classification, the finalized raster model encompassing seven classes was integrated with a comprehensive stack of vegetation indices and their respective statistical metrics, amounting to a total of 292 indicators. Numerous statistical metrics and tests are available for the assessment of the statistical significance pertaining to the differences observed between groups. In conducting a rigorous and thorough analysis to evaluate the statistical significance of various vegetation indices across distinct vegetation classes, we employed a variety of statistical techniques to ascertain the robustness and validity of our findings (Table S2-S8, Supplementary Materials)

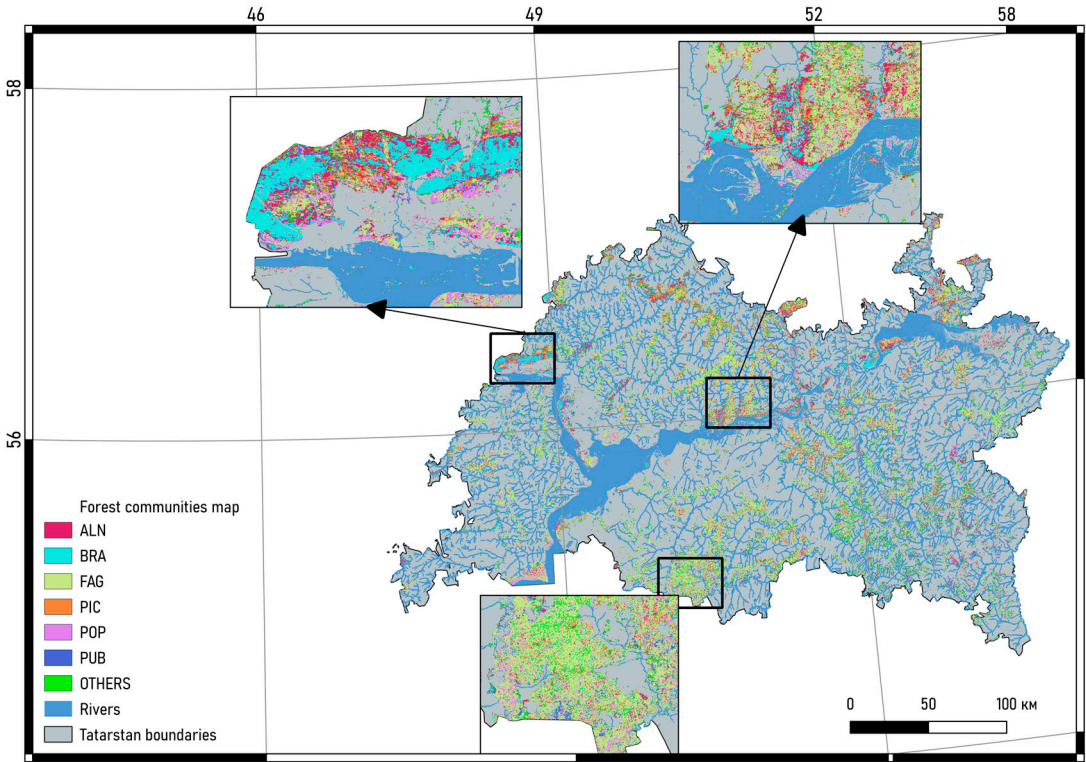


Figure 5. Result of spatial modelling of forest communities within the Republic of Tatarstan.

The dataset, characterized by a diverse range of vegetation classes, was subjected to a stringent preprocessing stage. The filtration of outliers was deemed paramount to ensure the accuracy and reliability of the results. A meticulous categorization of the vegetation indices was carried out into six specific groups: Reflectance-based indices, Water content indices, Soil-adjusted indices, Chlorophyll indices, Non-linear indices, and Structural indices. This categorization significantly

facilitated a nuanced and contextually relevant analysis, thereby enabling a more precise evaluation of the indices.

Utilizing the Kruskal-Wallis test, a non-parametric method, significant variances in the median values of the vegetation indices across different vegetation classes were identified. This preliminary step proved instrumental in pinpointing specific indices that merited further investigation. For a more detailed examination, a selection of statistical tests was employed, namely the t-test for comparing the mean values of two groups; the Kruskal-Wallis Test as a non-parametric alternative to ANOVA for three or more groups; Cohen's d effect size to gauge the magnitude of differences between two groups; and the Dunn Test as a post-hoc analysis following the Kruskal-Wallis Test. To account for Type I error during the tests, significance levels were adjusted using the Bonferroni correction. The findings of the statistical analysis are documented in the Table S9 at Supplementary Materials.

The meticulous approach towards statistical analysis entailed in this study was pivotal in delving deeper into the dynamics of vegetation indices across distinctive classes. The robust preprocessing of the dataset ensured a solid foundation for the subsequent statistical assessments. The strategic categorization of the vegetation indices provided a structured framework for a focused analysis, enhancing the interpretability and relevance of the findings.

Furthermore, the employment of both parametric (t-test) and non-parametric (Kruskal-Wallis Test and Dunn Test) statistical tests facilitated a well-rounded comprehension of the distribution and relationships among the vegetation indices across different vegetation classes. The utilization of Cohen's d effect size furnished a clearer understanding of the practical significance of the observed differences, complementing the p-value analysis from the t-test and Kruskal-Wallis Test.

Lastly, the implementation of the Bonferroni correction in the statistical analysis underscores a meticulous approach towards controlling the false positive rate, thereby fortifying the reliability of the findings. The detailed exposition of the statistical analysis outcomes in the supplementary materials ensures transparency and provides a solid foundation for further evaluation and interpretation of the results, significantly contributing to the broader understanding of vegetation indices dynamics across varying vegetation classes.

Indices that demonstrated significant disparities were then subjected to post-hoc analysis using the Mann-Whitney U tests for pairwise comparisons between vegetation classes. This granular approach enabled us to identify the precise pairs of vegetation classes that exhibited significant differences, providing invaluable insights. Furthermore, we calculated the median differences for each vegetation index across the vegetation classes. This provided a clear, straightforward metric to gauge the extent of the differences, adding another layer of understanding to our analysis.

By considering the p-values from the Kruskal-Wallis test, the outcomes of the post-hoc tests, and the median differences, we could deduce the most representative vegetation indices for each category. This comprehensive and holistic methodology ensured that our conclusions were not just statistically robust but also practically relevant.

However, it is imperative to note that statistical significance does not always equate to practical or ecological significance. The results must be interpreted within the specific research context, considering the practical applications of the vegetation indices. The outlier filtration process was crucial in enhancing data quality, subsequently bolstering the reliability of our findings. Additionally, when interpreting the results from the post-hoc tests, it was essential to adjust for multiple comparisons to mitigate the risk of false positives.

Incorporating a discussion perspective, this analysis sheds light on the intricate dynamics of vegetation indices across different vegetation classes. The identification of the most representative indices for each category stands as a testament to the importance of a thorough and contextually aware statistical approach. Similar methodologies have found applications in various ecological and remote sensing research endeavors, underscoring the universality and relevance of our approach.

Evaluating holistically, it is noteworthy that for all community groups, across all vegetation indices and statistical metrics, the highest level of significance was achieved, registering at 0.0 or

approaching it. This signifies the accuracy in the distribution of results from the unsupervised classification across community groups (Figure 6, Figures S1–S17 in supplementary materials).

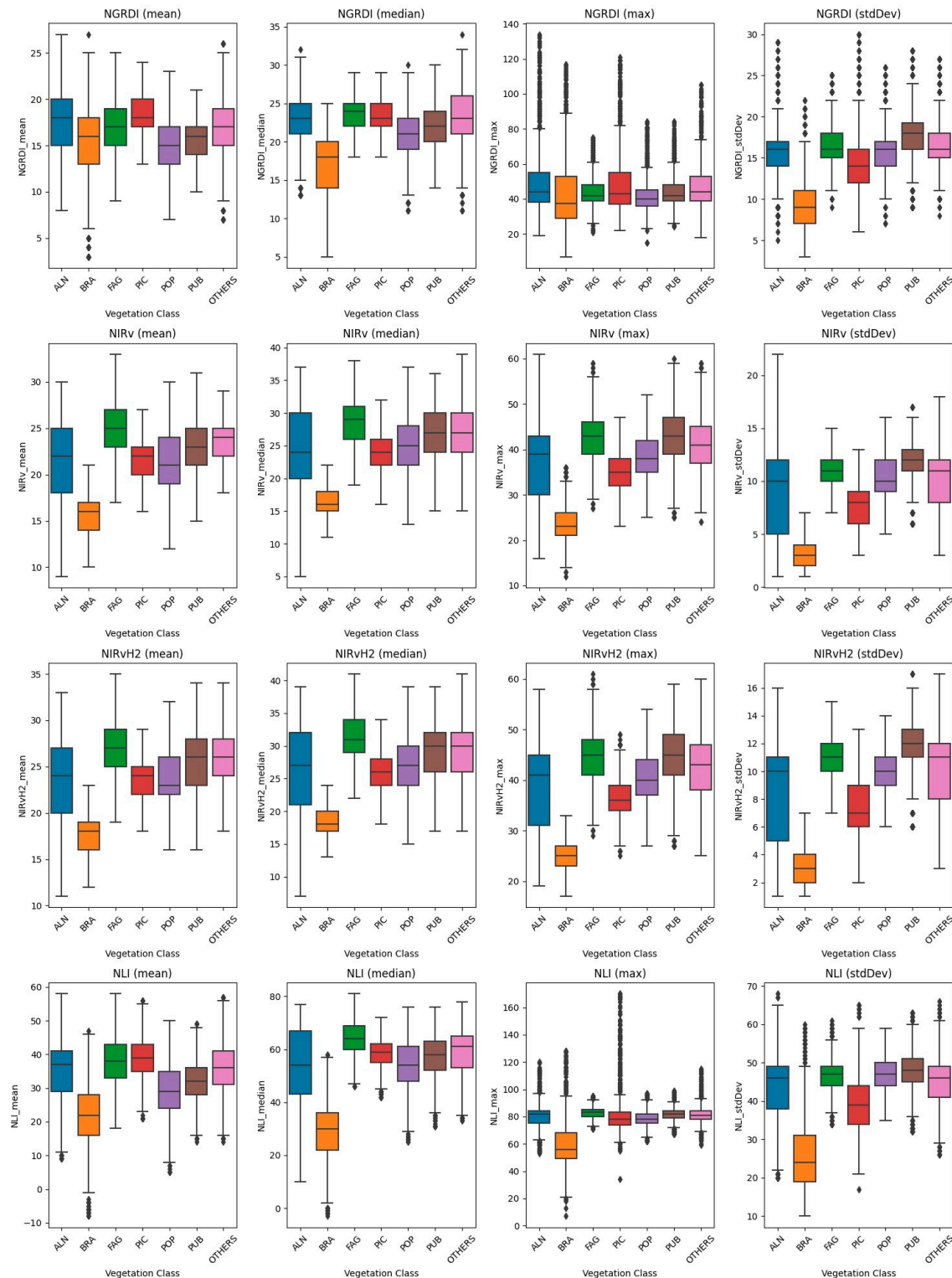


Figure 6. Distribution of statistical metrics of vegetation indices in forest community classes.

Within the selected groups of vegetation indices, one most significant vegetation index was chosen from each group post-testing, and the variations in value distribution within community groups were analyzed, as shown in Figure 7.

The detailed analysis of vegetation indices across various forest formations unveils distinct patterns, reflecting the intricate interplay between vegetation characteristics and environmental conditions. *Alnus*

(ALN), Fagus (FAG), and Pubescent Oak (PUB) formations consistently exhibit traits of robust and healthy vegetation, as evidenced by higher values in indices related to chlorophyll content, vegetation structure, and vitality. This can be attributed to their inherent biological characteristics and adaptation to their respective habitats, which might provide optimal conditions for growth and development.

In contrast, Brassica (BRA) and Picea (PIC) formations generally display lower values across the indices, suggesting less favorable conditions for vegetation. This could be a result of environmental stressors, less optimal soil conditions, or specific physiological adaptations of these species. The lower chlorophyll content and sparse vegetation structure might be indicative of these challenges.

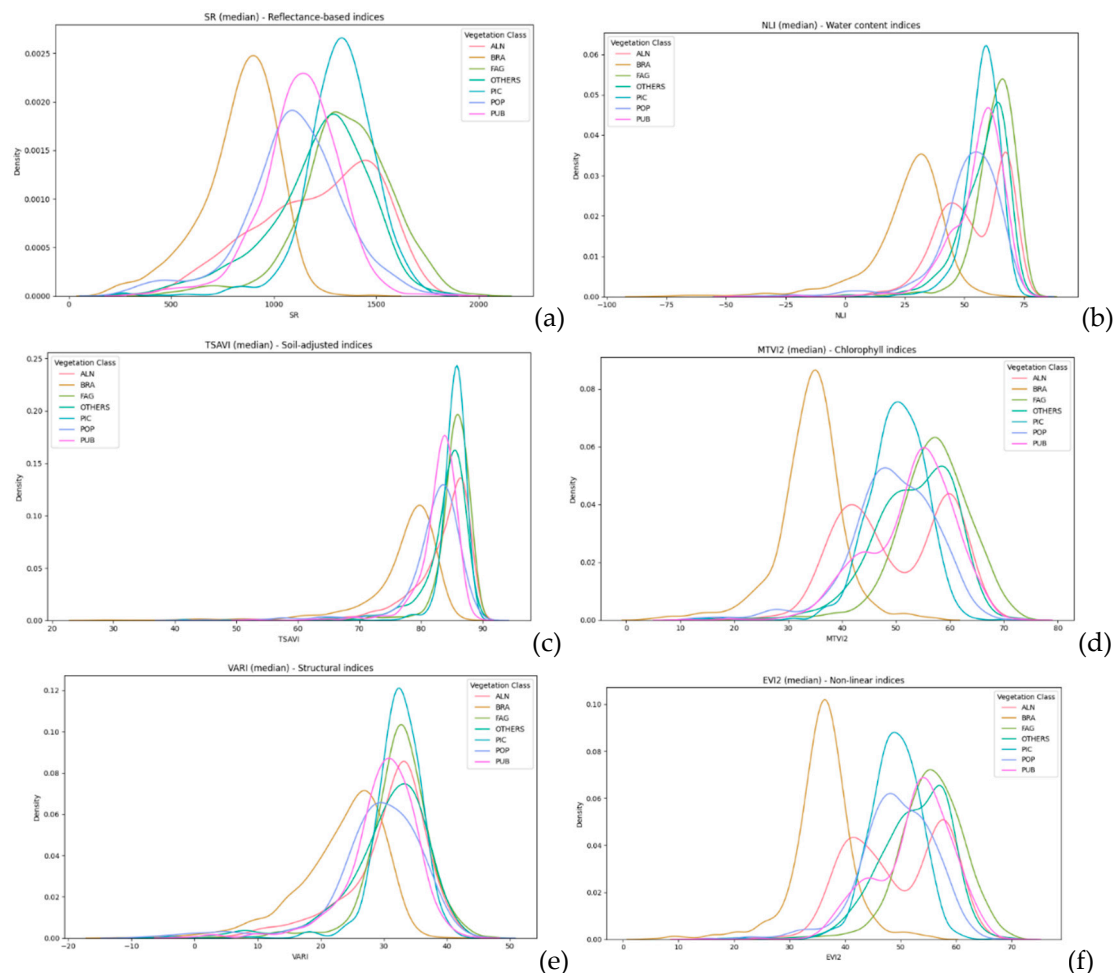


Figure 7. Variations in vegetation indices value distribution within community groups: reflectance-based indices (a), water content indices (b), soil-adjusted indices (c), chlorophyll indices (d), structural indices (e), non-linear indices (f).

Populus (POP) and OTHERS categories present a diverse range of conditions, as seen in the broad distribution across the indices. This diversity could be a reflection of the heterogeneity in species composition, age structures, and health conditions within these categories. The wide array of vegetation states captured in these categories underscores the complexity of forest ecosystems and the multitude of factors that influence vegetation dynamics.

Populus (POP), characterized by rapid growth and adaptability, tends to exhibit a broad range of values across the vegetation indices, indicating varying degrees of health and vitality within this formation. This variability could be attributed to the diverse age structures and health conditions of Populus species, as well as their ability to thrive in a variety of environmental conditions.

The OTHERS category, encompassing a mix of different species and vegetation types, showcases an even wider variability in the vegetation indices. This heterogeneity reflects the complexity of these

mixed formations, where multiple factors including species diversity, interspecies interactions, and environmental conditions come into play, influencing the overall vegetation health and structure.

The observed patterns in the vegetation indices across different forest formations provide valuable insights into the health and vitality of the vegetation, as well as the environmental conditions prevailing in these habitats. These patterns are closely aligned with the known biological characteristics and ecological preferences of the dominant species in each formation, offering a nuanced understanding of forest ecosystem dynamics.

The robustness of *Alnus*, *Fagus*, and Pubescent Oak formations can be linked to their ecological adaptability and resilience, which enable them to maintain high levels of chlorophyll content and structural integrity even in challenging conditions. On the other hand, the lower index values observed in *Brassica* and *Picea* formations highlight the potential challenges faced by these species, possibly due to environmental stressors or less optimal growth conditions.

These insights underscore the importance of continuous monitoring and analysis of vegetation indices for effective forest management and conservation efforts. By leveraging the nuanced information provided by these indices, we can gain a deeper understanding of forest ecosystem dynamics, identify areas in need of conservation or restoration, and implement targeted interventions to promote forest health and resilience.

4. Conclusions

The results of the study showed that through unsupervised classification of multispectral satellite images, areas of vegetation cover are identified that correspond with high confidence to vegetation classes in the Braun-Blanquet system. Based on the obtained model and data on the sequestration potential of various types of forest communities, future studies are planned to determine the carbon balance of ecosystems of the Republic of Tatarstan. These data can also serve as a cartographic base for planning and implementing climate projects.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Figure S1: Statistical features of unsupervised classification results across community groups; Table S1: Statistical analysis of correspondence between groups of raster values and specific classes of field data; Table S2-S8: Statistics on vegetation metric across community groups; Table S9: Statistical verification of classification results.

Author Contributions: “Conceptualization, V.P. and M.K.; methodology, A.G.; software, A.G.; validation, A.G., V.P.; formal analysis, A.G.; investigation, V.P., M.K.; writing—original draft preparation, A.G.; writing—review and editing, B.U.; visualization, A.G., B.U.; supervision, V.P.; project administration, M.K.; funding acquisition, B.U. All authors have read and agreed to the published version of the manuscript.”

Funding: This work was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, project № FZSM-2022-0003.

Data Availability Statement: Resulted vegetation metrics on 7 forest communities classes represented at Supplementary Materials Table S2-S8.

Conflicts of Interest: “The authors declare no conflict of interest.”

References

1. Kurbatova, A.I. Analytical Review of Modern Studies of Changes in the Biotic Components of the Carbon Cycle. *Vestn. Ross. univ. družby nar., Ser. Èkol. bezop. žiznedeât.* **2020**, *28*, 428–438, doi:10.22363/2313-2310-2020-28-4-428-438.
2. Mukharamova, S.; Saveliev, A.; Ivanov, M.; Gafurov, A.; Yermolaev, O. Estimating the Soil Erosion Cover-Management Factor at the European Part of Russia. *IJGI* **2021**, *10*, 645, doi:10.3390/ijgi10100645.
3. Gafurov, A. Mapping of Rill Erosion of the Middle Volga (Russia) Region Using Deep Neural Network. *ISPRS International Journal of Geo-Information* **2022**, *11*, 197, doi:10.3390/ijgi11030197.
4. Fardeeva, M.; Lukyanova, Y.; Eskina, A.; Usmanov, B. Distribution and Habitat Features of Rare Orchid Species (*Orchidaceae* Juss.) in the National Park “Nizhnaya Kama.” *E3S Web Conf.* **2023**, *420*, 01023, doi:10.1051/e3sconf/202342001023.

5. Executive Steering Committee for Australian Vegetation Information *Australian Vegetation Attribute Manual: National Vegetation Information System, Version 6.0*; Department of Environment and Heritage: Canberra, 2003;
6. Faber-Langendoen, D.; Keeler-Wolf, T.; Meidinger, D.; Tart, D.; Hoagland, B.; Josse, C.; Navarro, G.; Ponomarenko, S.; Saucier, J.-P.; Weakley, A.; et al. EcoVeg: A New Approach to Vegetation Description and Classification. *Ecological Monographs* **2014**, *84*, 533–561, doi:10.1890/13-2334.1.
7. De Cáceres, M.; Chytrý, M.; Agrillo, E.; Attorre, F.; Botta-Dukát, Z.; Capelo, J.; Czúcz, B.; Dengler, J.; Ewald, J.; Faber-Langendoen, D.; et al. A Comparative Framework for Broad-scale Plot-based Vegetation Classification. *Applied Vegetation Science* **2015**, *18*, 543–560, doi:10.1111/avsc.12179.
8. Kozhevnikova, M.; Prokhorov, V. Syntaxonomy of the Xero-Mesophytic Oak Forests in the Republic of Tatarstan (Eastern Europe). *VCS* **2021**, *2*, 47–58, doi:10.3897/VCS/2021/39583.
9. Mucina, L.; Bültmann, H.; Dierßen, K.; Theurillat, J.; Raus, T.; Čarni, A.; Šumberová, K.; Willner, W.; Dengler, J.; García, R.G.; et al. Vegetation of Europe: Hierarchical Floristic Classification System of Vascular Plant, Bryophyte, Lichen, and Algal Communities. *Applied Vegetation Science* **2016**, *19*, 3–264, doi:10.1111/avsc.12257.
10. De Cáceres, M.; Font, X.; Oliva, F. The Management of Vegetation Classifications with Fuzzy Clustering: Fuzzy Clustering in Vegetation Classifications. *Journal of Vegetation Science* **2010**, *21*, 1138–1151, doi:10.1111/j.1654-1103.2010.01211.x.
11. Gafurov, A. The Methodological Aspects of Constructing a High-Resolution DEM of Large Territories Using Low-Cost UAVs on the Example of the Sarycum Aeolian Complex, Dagestan, Russia. *Drones* **2021**, *5*, 7, doi:10.3390/drones5010007.
12. Duda, T.; Canty, M. Unsupervised Classification of Satellite Imagery: Choosing a Good Algorithm. *International Journal of Remote Sensing* **2002**, *23*, 2193–2212, doi:10.1080/01431160110078467.
13. Liu, X. Supervised Classification and Unsupervised Classification.; ATS, 2005.
14. Kozak, M.; Scaman, C.H. Unsupervised Classification Methods in Food Sciences: Discussion and Outlook. *J Sci Food Agric* **2008**, *88*, 1115–1127, doi:10.1002/jsfa.3215.
15. Bandyopadhyay, S.; Saha, S. *Unsupervised Classification: Similarity Measures, Classical and Metaheuristic Approaches, and Applications*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; ISBN 978-3-642-32450-5.
16. Olaode, A.; Naghdy, G.; Todd, C. Unsupervised Classification of Images: A Review. *International Journal of Image Processing* **2014**, *8*, 325–342.
17. Xie, Y.; Sha, Z.; Yu, M. Remote Sensing Imagery in Vegetation Mapping: A Review. *Journal of Plant Ecology* **2008**, *1*, 9–23, doi:10.1093/jpe/rtm005.
18. Ma, Z.; Liu, Z.; Zhao, Y.; Zhang, L.; Liu, D.; Ren, T.; Zhang, X.; Li, S. An Unsupervised Crop Classification Method Based on Principal Components Isometric Binning. *ISPRS International Journal of Geo-Information* **2020**, *9*, 648, doi:10.3390/ijgi9110648.
19. Anchang, J.Y.; Ananga, E.O.; Pu, R. An Efficient Unsupervised Index Based Approach for Mapping Urban Vegetation from IKONOS Imagery. *International Journal of Applied Earth Observation and Geoinformation* **2016**, *50*, 211–220, doi:10.1016/j.jag.2016.04.001.
20. Ragetti, S.; Herberz, T.; Siegfried, T. An Unsupervised Classification Algorithm for Multi-Temporal Irrigated Area Mapping in Central Asia. *Remote Sensing* **2018**, *10*, 1823, doi:10.3390/rs10111823.
21. Yermolaev, O.; Usmanov, B.; Gafurov, A.; Poesen, J.; Vedeneva, E.; Lisetskii, F.; Nicu, I.C. Assessment of Shoreline Transformation Rates and Landslide Monitoring on the Bank of Kuibyshev Reservoir (Russia) Using Multi-Source Data. *Remote Sensing* **2021**, *13*, 4214, doi:10.3390/rs13214214.
22. Perevedentsev, Y.P.; Shantalinskii, K.M.; Guryanov, V.V.; Aukhadeev, T.R. Climatic Changes on the Territory of the Volga Federal District. *IOP Conf. Ser.: Earth Environ. Sci.* **2020**, *606*, 012045, doi:10.1088/1755-1315/606/1/012045.
23. Olson, D.M.; Dinerstein, E.; Wikramanayake, E.D.; Burgess, N.D.; Powell, G.V.N.; Underwood, E.C.; D'amico, J.A.; Itoua, I.; Strand, H.E.; Morrison, J.C.; et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* **2001**, *51*, 933, doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.
24. Yermolaev, O.; Usmanov, B. The Basin Approach to the Anthropogenic Impact Assessment in Oil-Producing Region. In Proceedings of the International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM; Bulgaria, June 17 2014; Vol. 2, pp. 681–688.
25. Prokhorov, V.; Rogova, T.; Kozhevnikova, M. Vegetation Database of Tatarstan. *phyto* **2017**, *47*, 309–313, doi:10.1127/phyto/2017/0172.

26. Gafurov, A.; Gainullin, I.; Usmanov, B.; Khomyakov, P.; Kasimov, A. Impacts of Fluvial Processes on Medieval Settlement Lukovskoe (Tatarstan, Russia). *Proc. IAHS* **2019**, *381*, 31–35, doi:10.5194/piahs-381-31-2019.
27. Usmanov, B.; Gainullin, I.; Gafurov, A.; Ivanov, M.; Khomyakov, P.; Gubaidullin, A.; Nicu, I.C. Web-Portal “Country of Cities” - Comprehensive Study of the Volga Bulgarian Fortified Settlements. *International Journal of Conservation Science* **2022**, *13*, 1143–1158.
28. Landsat Science Available online: <https://landsat.gsfc.nasa.gov/> (accessed on 7 March 2020).
29. Montero, D. Eemont: A Python Package That Extends Google Earth Engine. *JOSS* **2021**, *6*, 3168, doi:10.21105/joss.03168.
30. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sensing of Environment* **2017**, *202*, 18–27, doi:10.1016/j.rse.2017.06.031.
31. Google Earth Engine Available online: <https://earthengine.google.com> (accessed on 26 May 2021).
32. Roujean, J.-L.; Breon, F.-M. Estimating PAR Absorbed by Vegetation from Bidirectional Reflectance Measurements. *Remote Sensing of Environment* **1995**, *51*, 375–384, doi:10.1016/0034-4257(94)00114-3.
33. Yang, W.; Kobayashi, H.; Wang, C.; Shen, M.; Chen, J.; Matsushita, B.; Tang, Y.; Kim, Y.; Bret-Harte, M.S.; Zona, D.; et al. A Semi-Analytical Snow-Free Vegetation Index for Improving Estimation of Plant Phenology in Tundra and Grassland Ecosystems. *Remote Sensing of Environment* **2019**, *228*, 31–44, doi:10.1016/j.rse.2019.03.028.
34. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sensing of Environment* **1996**, *58*, 289–298, doi:10.1016/S0034-4257(96)00072-7.
35. Wang, F.; Huang, J.; Tang, Y.; Wang, X. New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. *Rice Science* **2007**, *14*, 195–203, doi:10.1016/S1672-6308(07)60027-4.
36. Jurgens, C. The Modified Normalized Difference Vegetation Index (mNDVI) a New Index to Determine Frost Damages in Agriculture Based on Landsat TM Data. *International Journal of Remote Sensing* **1997**, *18*, 3583–3594, doi:10.1080/014311697216810.
37. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS.; January 1 1974.
38. Bannari, A.; Asalhi, H.; Teillet, P.M. Transformed Difference Vegetation Index (TDVI) for Vegetation Cover Mapping. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium; IEEE: Toronto, Ont., Canada, 2002; Vol. 5, pp. 3053–3055.
39. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology* **2004**, *161*, 165–173, doi:10.1078/0176-1617-01176.
40. Clevers, J.G.P.W. Application of a Weighted Infrared-Red Vegetation Index for Estimating Leaf Area Index by Correcting for Soil Moisture. *Remote Sensing of Environment* **1989**, *29*, 25–37, doi:10.1016/0034-4257(89)90076-X.
41. Apan, A.; Held, A.; Phinn, S.; Markley, J. Detecting Sugarcane ‘Orange Rust’ Disease Using EO-1 Hyperion Hyperspectral Imagery. *International Journal of Remote Sensing* **2004**, *25*, 489–498, doi:10.1080/01431160310001618031.
42. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Weisz, R. Aerial Color Infrared Photography for Determining Late-Season Nitrogen Requirements in Corn. *Agronomy Journal* **2005**, *97*, 1443–1451, doi:10.2134/agronj2004.0314.
43. Badgley, G.; Field, C.B.; Berry, J.A. Canopy Near-Infrared Reflectance and Terrestrial Photosynthesis. *Sci. Adv.* **2017**, *3*, e1602244, doi:10.1126/sciadv.1602244.
44. Zeng, Y.; Hao, D.; Badgley, G.; Damm, A.; Rascher, U.; Ryu, Y.; Johnson, J.; Krieger, V.; Wu, S.; Qiu, H.; et al. Estimating Near-Infrared Reflectance of Vegetation from Hyperspectral Data. *Remote Sensing of Environment* **2021**, *267*, 112723, doi:10.1016/j.rse.2021.112723.
45. Klemas, V.; Smart, R. The Influence of Soil Salinity, Growth Form, and Leaf Moisture on-the Spectral Radiance Of. *Photogramm. Eng. Remote Sens* **1983**, *49*, 77–83.
46. Wilson, E.H.; Sader, S.A. Detection of Forest Harvest Type Using Multiple Dates of Landsat TM Imagery. *Remote Sensing of Environment* **2002**, *80*, 385–396, doi:10.1016/S0034-4257(01)00318-2.
47. Wang, L.; Qu, J.J. NMDI: A Normalized Multi-Band Drought Index for Monitoring Soil and Vegetation Moisture with Satellite Remote Sensing. *Geophys. Res. Lett.* **2007**, *34*, L20405, doi:10.1029/2007GL031021.

48. Huntjr, E.; Rock, B. Detection of Changes in Leaf Water Content Using Near- and Middle-Infrared Reflectances☆. *Remote Sensing of Environment* **1989**, *30*, 43–54, doi:10.1016/0034-4257(89)90046-1.
49. Wu, W. The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization. *Remote Sensing* **2014**, *6*, 1211–1233, doi:10.3390/rs6021211.
50. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sensing of Environment* **1994**, *48*, 119–126, doi:10.1016/0034-4257(94)90134-1.
51. Rondeaux, G.; Steven, M.; Baret, F. Optimization of Soil-Adjusted Vegetation Indices. *Remote Sensing of Environment* **1996**, *55*, 95–107, doi:10.1016/0034-4257(95)00186-7.
52. Kaufman, Y.J.; Tanre, D. Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sensing* **1992**, *30*, 261–270, doi:10.1109/36.134076.
53. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment* **1988**, *25*, 295–309, doi:10.1016/0034-4257(88)90106-X.
54. Baret, F.; Guyot, G.; Major, D.J. TSAVI: A Vegetation Index Which Minimizes Soil Brightness Effects On LAI And APAR Estimation. In Proceedings of the 12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium, IEEE: Vancouver, Canada, 1989; Vol. 3, pp. 1355–1358.
55. Gillespie, A.R.; Kahle, A.B.; Walker, R.E. Color Enhancement of Highly Correlated Images. II. Channel Ratio and “Chromaticity” Transformation Techniques. *Remote Sensing of Environment* **1987**, *22*, 343–365, doi:10.1016/0034-4257(87)90088-5.
56. Bendig, J.; Yu, K.; Aasen, H.; Bolten, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.L.; Bareth, G. Combining UAV-Based Plant Height from Crop Surface Models, Visible, and near Infrared Vegetation Indices for Biomass Monitoring in Barley. *International Journal of Applied Earth Observation and Geoinformation* **2015**, *39*, 79–87, doi:10.1016/j.jag.2015.02.012.
57. Haboudane, D. Hyperspectral Vegetation Indices and Novel Algorithms for Predicting Green LAI of Crop Canopies: Modeling and Validation in the Context of Precision Agriculture. *Remote Sensing of Environment* **2004**, *90*, 337–352, doi:10.1016/j.rse.2003.12.013.
58. Chen, J.M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Canadian Journal of Remote Sensing* **1996**, *22*, 229–242, doi:10.1080/07038992.1996.10855178.
59. Vincini, M.; Frazzi, E.; D'Alessio, P. A Broad-Band Leaf Chlorophyll Vegetation Index at the Canopy Scale. *Precision Agric* **2008**, *9*, 303–319, doi:10.1007/s11119-008-9075-z.
60. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel Algorithms for Remote Estimation of Vegetation Fraction. *Remote Sensing of Environment* **2002**, *80*, 76–87, doi:10.1016/S0034-4257(01)00289-9.
61. Vincini, M.; Frazzi, E. Comparing Narrow and Broad-Band Vegetation Indices to Estimate Leaf Chlorophyll Content in Planophile Crop Canopies. *Precision Agric* **2011**, *12*, 334–344, doi:10.1007/s11119-010-9204-3.
62. John E. Plnder III; McLeod, K.W. Indications of Relative Drought Stress in Longleaf Pine from Thematic Mapper Data. *Photogrammetric Engineering & Remote Sensing* **1999**, *65*, 495–501.
63. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666, doi:10.2307/1936256.
64. Jiang, Z.; Huete, A.; Didan, K.; Miura, T. Development of a Two-Band Enhanced Vegetation Index without a Blue Band. *Remote Sensing of Environment* **2008**, *112*, 3833–3845, doi:10.1016/j.rse.2008.06.006.
65. Louhaichi, M.; Borman, M.M.; Johnson, D.E. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. *Geocarto International* **2001**, *16*, 65–70, doi:10.1080/10106040108542184.
66. Pinty, B.; Verstraete, M.M. GEMI: A Non-Linear Index to Monitor Global Vegetation from Satellites. *Vegetatio* **1992**, *101*, 15–20, doi:10.1007/BF00031911.
67. Escadafal, R.; Huete, A. Etude Des Propriétés Spectrales Des Sols Arides Appliquée à l'amélioration Des Indices de Végétation Obtenus Par Télédétection. *Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre* **1991**, *312*, 1385–1391.
68. Rikimaru, A.; Roy, P.S.; Miyatake, S. Tropical Forest Cover Density Mapping. *Tropical ecology* **2002**, *43*, 39–47.
69. Jiang, H.; Wang, S.; Cao, X.; Yang, C.; Zhang, Z.; Wang, X. A Shadow- Eliminated Vegetation Index (SEVI) for Removal of Self and Cast Shadow Effects on Vegetation in Rugged Terrains. *International Journal of Digital Earth* **2019**, *12*, 1013–1029, doi:10.1080/17538947.2018.1495770.
70. Hancock, D.W.; Dougherty, C.T. Relationships between Blue- and Red-based Vegetation Indices and Leaf Area and Yield of Alfalfa. *Crop Science* **2007**, *47*, 2547–2556, doi:10.2135/cropsci2007.01.0031.

71. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. *Journal of Plant Physiology* **2003**, *160*, 271–282, doi:10.1078/0176-1617-00887.
72. Yang, P.; Van Der Tol, C.; Campbell, P.K.E.; Middleton, E.M. Fluorescence Correction Vegetation Index (FCVI): A Physically Based Reflectance Index to Separate Physiological and Non-Physiological Information in Far-Red Sun-Induced Chlorophyll Fluorescence. *Remote Sensing of Environment* **2020**, *240*, 111676, doi:10.1016/j.rse.2020.111676.
73. Ren-hua, Z.; N, R.N.X. and L.K. Approach for a Vegetation Index Resistant to Atmospheric Effect. *Journal of Integrative Plant Biology* **1996**, *38*.
74. Kawashima, S. An Algorithm for Estimating Chlorophyll Content in Leaves Using a Video Camera. *Annals of Botany* **1998**, *81*, 49–54, doi:10.1006/anbo.1997.0544.
75. Crippen, R. Calculating the Vegetation Index Faster. *Remote Sensing of Environment* **1990**, *34*, 71–73, doi:10.1016/0034-4257(90)90085-Z.
76. Guo, Y.; Wang, H.; Wu, Z.; Wang, S.; Sun, H.; Senthilnath, J.; Wang, J.; Robin Bryant, C.; Fu, Y. Modified Red Blue Vegetation Index for Chlorophyll Estimation and Yield Prediction of Maize from Visible Images Captured by UAV. *Sensors* **2020**, *20*, 5055, doi:10.3390/s20185055.
77. Peng Gong; Ruiliang Pu; Biging, G.S.; Larrieu, M.R. Estimation of Forest Leaf Area Index Using Vegetation Indices Derived from Hyperion Hyperspectral Data. *IEEE Trans. Geosci. Remote Sensing* **2003**, *41*, 1355–1362, doi:10.1109/TGRS.2003.812910.
78. Gu, Y.; Brown, J.F.; Verdin, J.P.; Wardlow, B. A Five-Year Analysis of MODIS NDVI and NDWI for Grassland Drought Assessment over the Central Great Plains of the United States. *Geophys. Res. Lett.* **2007**, *34*, L06407, doi:10.1029/2006GL029127.
79. Wang, C.; Chen, J.; Wu, J.; Tang, Y.; Shi, P.; Black, T.A.; Zhu, K. A Snow-Free Vegetation Index for Improved Monitoring of Vegetation Spring Green-up Date in Deciduous Ecosystems. *Remote Sensing of Environment* **2017**, *196*, 1–12, doi:10.1016/j.rse.2017.04.031.
80. Sulik, J.J.; Long, D.S. Spectral Considerations for Modeling Yield of Canola. *Remote Sensing of Environment* **2016**, *184*, 161–174, doi:10.1016/j.rse.2016.06.016.
81. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment* **1979**, *8*, 127–150, doi:10.1016/0034-4257(79)90013-0.
82. Han, J.; Zhang, Z.; Cao, J. Developing a New Method to Identify Flowering Dynamics of Rapeseed Using Landsat 8 and Sentinel-1/2. *Remote Sensing* **2020**, *13*, 105, doi:10.3390/rs13010105.
83. Saberioon, M.M.; Amin, M.S.M.; Anuar, A.R.; Gholizadeh, A.; Wayayok, A.; Khairunniza-Bejo, S. Assessment of Rice Leaf Chlorophyll Content Using Visible Bands at Different Growth Stages at Both the Leaf and Canopy Scale. *International Journal of Applied Earth Observation and Geoinformation* **2014**, *32*, 35–45, doi:10.1016/j.jag.2014.03.018.
84. Hunt, E.R.; Doraiswamy, P.C.; McMurtrey, J.E.; Daughtry, C.S.T.; Perry, E.M.; Akhmedov, B. A Visible Band Index for Remote Sensing Leaf Chlorophyll Content at the Canopy Scale. *International Journal of Applied Earth Observation and Geoinformation* **2013**, *21*, 103–112, doi:10.1016/j.jag.2012.07.020.
85. Broge, N.H.; Leblanc, E. Comparing Prediction Power and Stability of Broadband and Hyperspectral Vegetation Indices for Estimation of Green Leaf Area Index and Canopy Chlorophyll Density. *Remote Sensing of Environment* **2001**, *76*, 156–172, doi:10.1016/S0034-4257(00)00197-8.
86. Hansen, M.C.; Potapov, P.V.; Pickens, A.H.; Tyukavina, A.; Hernandez-Serna, A.; Zalles, V.; Turubanova, S.; Kommareddy, I.; Stehman, S.V.; Song, X.-P.; et al. Global Land Use Extent and Dispersion within Natural Land Cover Using Landsat Data. *Environ. Res. Lett.* **2022**, *17*, 034050, doi:10.1088/1748-9326/ac46ec.
87. Kozhevnikova, M.V.; Prokhorov, V.E.; Rogova, T.V. Xeromesophytic Broad-Leaved Forest Communities of the Republic of Tatarstan in the Hierarchy of Syntaxa within the Braun-Blanquet System. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Estestvennye Nauki* **2018**, *160*, 445–458.
88. Kazan Federal University (Kazan, Russian Federation); Kozhevnikova, M.V.; Prokhorov, V.E.; Kazan Federal University (Kazan, Russian Federation); Saveliev, A.A.; Kazan Federal University (Kazan, Russian Federation) Predictive Modeling for the Distribution of Plant Communities of the Order Quercetalia Pubescenti-Petraeae Klika 1933. *Tomsk State University Journal of Biology* **2019**, *59*–73, doi:10.17223/19988591/47/4.
89. Tichý, L. JUICE, Software for Vegetation Classification. *J Vegetation Science* **2002**, *13*, 451–453, doi:10.1111/j.1654-1103.2002.tb02069.x.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.