University of Tartu

Faculty of Science and Technology

Institute of Ecology and Earth Sciences

Department of Geography

Master's thesis in Geoinformatics for Urbanised Society (30 ECTS)

Comparing Raster Visualization Techniques for Environmental Indices: Univariate and Bivariate Mapping in Estonia

Wenyi Fang

Supervisors:

Alexander Kmoch, PhD

Tartu 2025

Abstract

Comparing raster visualization techniques for environmental indices: univariate and bivariate mapping in Estonia

Visualizing environmental data effectively is crucial for understanding ecological patterns and communicating spatial information. This study compares three cartographic visualization techniques for raster data: univariate mapping, bivariate mapping, and value-by-alpha mapping. To achieve this, this research applied Principal Component Analysis (PCA) to five remote sensing indices (NDVI, NDWI, BSI, NDMI, and LST), and visualized the first two principal components (PC1 and PC2) across four diverse landscapes in Estonia. The objective is to assess the strengths and limitations of these methods in representing complex ecological conditions.

Univariate maps were created using a GIS-based weighted overlay method, applying PC1 loadings to generate single-variable sustainability maps. Bivariate mapping included standard bivariate maps, corner models emphasizing extreme combinations, and diagonal models highlighting variable interactions. Value-by-alpha maps employed color to encode PC1 and transparency for PC2, enhancing interpretability. At the same time, quantile and equal interval classifications were applied and compared to illustrate their influence on visual contrast and interpretation.

Results indicate that univariate maps offer clear, easily interpretable spatial distributions which are suitable for public communication. Bivariate maps effectively display complex ecological states, with the corner model highlighting extremes and the diagonal model emphasizing variable divergence. Value-by-alpha maps selectively highlight critical zones, though their performance depends on transparency settings. Quantile classification enhances contrast but can overemphasize rare values, while equal interval classification maintains a direct representation of data ranges. Ultimately, the choice of classification and visualization method must align with map objectives whether for exploration analysis, public communication, or expert interpretation.

The proposed framework advances multivariate environmental visualization by integrating PCA-based dimensionality reduction with tailored cartographic strategies, offering practical guidance for effective ecological communication and decision-making.

Keywords: cartographic techniques, bivariate map, raster visualization

CERCS code: P510 – cartography, T181 – Remote sensing

Annotatsioon

Rastripõhiste visualiseerimistehnikate võrdlus keskkonnaindeksite jaoks: ühemõõtmeline ja kahemõõtmeline kaardistamine Eestis

Keskkonnaandmete tõhus visualiseerimine on oluline ökoloogiliste mustrite mõistmiseks ja ruumilise teabe edastamiseks. Käesolev töö võrdleb kolme kartograafilist visualiseerimistehnikat: ühetunnuselist kaardistamist (univariate mapping), kahetunnuselist kaardistamist (bivariate mapping) ja läbipaistvusepõhist kaardistamist (value-by-alpha mapping). Selle saavutamiseks rakendati kaugseire indeksitele (NDVI, NDWI, BSI, NDMI ja LST) peamiste komponentide analüüsi (PCA), ning visualiseeriti kaks esimest põhikomponenti (PC1 ja PC2) neljal erineval maastikul Eestis. Eesmärk on hinnata nende meetodite tugevusi ja piiranguid keerukate ökoloogiliste tingimuste kujutamisel.

Ühemõõtmelised kaardid loodi GIS-põhise kaalutud ülekatte meetodi abil, kasutades PC1 koormusi, et luua ühetunnuselisi jätkusuutlikkuse kaarte. Kahemõõtmelised kaardid hõlmasid standardseid kahetunnuselisi kaarte, nurkmudeleid, mis tõid välja äärmuslikke kombinatsioone, ja diagonaalmudeleid, mis rõhutasid muutujate vahelisi interaktsioone. Läbipaistvuse põhised kaardid kasutasid arusaadavuse parandamiseks PC1 kodeerimisel värve ja PC2 kodeerimisel läbipaistvust. Samal ajal rakendati kvantiil- ja võrdsete intervallide klassifikatsioone ning võrreldi nende mõju visuaalsele kontrastile ja tõlgendusele.

Tulemused näitavad, et ühemõõtmelised kaardid on selge ja kergesti mõistetava ruumilise jaotusega, mis sobivad avalikuks suhtluseks. Kahemõõtmelised kaardid võimaldavad tõhusalt kujutada erinevaid ökoloogilisi aspekte: nurkmudel äärmuseid ja diagonaalmudel muutujate lahknevust. Läbipaistvuse põhised kaardid toovad esile kriitilised tsoonid, kuid on tundlikud läbipaistvusseadete suhtes. Kvantiilklassifikatsioon suurendab kontrasti, kuid võib üle tähtsustada haruldasi väärtusi. Võrdsete intervallide klassifikatsioon aga säilitab andmevahemike otsese esitluse. Kaardi eesmärkidest lähtuvalt, näiteks andmete uurimine, avalik suhtlus või eksperttõlgenduseks, tuleb valida sobiv klassifitseerimis- ja visualiseerimismeetod.

Käesolev uurimistöö edendab mitmemõõtmelist keskkonnaalast visualiseerimist, ühendades PCA-põhise dimensioonide vähendamise kohandatud kartograafiliste strateegiatega ning pakkudes praktilisi juhiseid tõhusaks ökoloogiliseks kommunikatsiooniks ja otsuste tegemiseks.

Märksõnad: kartograafilised tehnikad, kahemõõtmeline kaart, rastervisualiseerimine

CERCS code: P510 – kartograafia, T181 – kaugseire

Table of Contents

Introduction

Industrialization, urbanization, and the expansion of agriculture have significantly altered natural ecosystems, resulting in widespread land cover changes and increasing environmental pressures. Environmental pollution and waste discharge cause global warming and climate anomalies (X. Zhang et al., 2024). In this context, the ability to effectively communicate ecological conditions and spatial patterns is essential for supporting environmental awareness, monitoring, and decision-making (Pettorelli et al., 2005). Remote sensing technology, with its large-scale and dynamic observations, plays a significant role in capturing surface conditions for such communication purposes (J. Li et al., 2020).

Given the multidimensional nature of ecological systems, capturing environmental conditions requires integrating information about vegetation, water content, soil exposure, and surface temperature. Relying on a single indicator often fails to represent the full complexity of landscape conditions, particularly in heterogeneous or human-impacted regions. Remote sensing offers a wide range of spectral indices that capture different ecological dimensions (Hu & Xu, 2018). In ecological research, the Normalized Difference Vegetation Index (NDVI) is not only used for vegetation detection but also for the analysis of climate change, human activities and environmental events (Teffera et al., 2018). The Normalized Difference Water Index (NDWI) is commonly used to assess the water content in vegetation and identify water bodies (Ferreira et al., 2024). Other indices, such as the Normalized Difference Moisture Index (NDMI), Land Surface Temperature (LST), and the Bare Soil Index (BSI), provide complementary insights into surface conditions, moisture stress, and soil exposure (Panahi et al., 2024; Sathyaseelan et al., 2023).

Understanding the relationship and changing patterns of soil, vegetation and water bodies through these indicators is fundamental for identifying areas of environmental stress, monitoring ecological degradation, and guiding conservation efforts. Remote sensing indices enable these components to be captured at broad scales, offering valuable insights into ecological conditions. However, due to the multi-dimensional nature of these indices and their complex interrelationships, effectively analyzing and visualizing them remains a challenge (Yagoub et al., 2022). Traditional methods rely on univariate color schemes, which often fail to convey subtle relationships or anomalies, such as vegetation stress in non-arid areas or soil degradation masked by surface indicators. Novel visualization techniques are essential for transforming multi-index environmental data into intuitive and interpretable

spatial representations that support environmental monitoring and policy formulation. In this context, data-driven methods such as Principal Component Analysis (PCA) combined with cartographic techniques that use RGB color encoding and alpha transparency, offer effective solutions for visualizing multivariate spatial patterns.

PCA is often used to reduce the dimensionality of multispectral satellite images and minimize redundancy among highly correlated indices (Faisal & Shaker, 2017; Palanisamy et al., 2023). In this study, PCA is applied to five remote sensing indices: NDVI, NDWI, LST, NDMI, and BSI, selected for their ecological relevance and temporal consistency. This transformation extracts the most informative environmental gradients while simplifying the complex multivariate structure of raster data.

In Estonia, for instance, most public web maps rely on vector representations and do not incorporate raster-based environmental indicators. There is a gap in the integration of spatial analysis and dynamic cartography for raster data, limiting public access to critical environmental insights.

To address these challenges, this thesis develops a PCA-based visualization framework and test it across four contrasting regions in Estonia: Tallinn, Tartu, the Porijõgi catchment, and the northeastern Ida-Viru Mining Region, which represent diverse landscape types and ecological gradients. Three visualization strategies are employed to visualize PCA-reduced components: univariate index mapping, bivariate color compositions, and value-by-alpha mapping. Together, these methods aim to improve the clarity and interpretability of environmental data and support scalable mapping practices.

This research is guided by two main questions:

1. How can PCA-based dimensionality reduction enhance the effectiveness of multivariate environmental raster visualizations?
2. How can the visual expression of multivariate raster indices be improved to support effective communication of environmental information?

1 Theoretical Overview

1.1 Multivariate Integration of Remote Sensing Indices

Numerous studies have employed remote sensing techniques using single environmental indicators, such as vegetation indices or land surface temperature, to assess the ecological condition of forests, wetlands, and urban areas. However, single-variable assessments fall short in capturing the complexity of ecosystems and may neglect the contributions of other key ecological factors (T. Zhang et al., 2021). For a more comprehensive evaluation of ecological quality, it is essential to integrate multiple environmental indices. Various composite indices have been developed to serve this purpose. For example, Messer et al. (2014) developed the Environmental Quality Index (EQI) which is based on the county level in the United States, focuses on the relationship between environmental factors and human health. Similarly, Shah et al. (2019) proposed the Energy Security and Environmental Sustainability Index to measure the intersection of energy concerns and ecological sustainability. However, these indices typically rely on socioeconomic or field-based data, making them complex and not fully compatible with remote sensing data streams.

A common approach in remote sensing for multivariate integration is the GIS overlay technique, which aggregates multiple univariate thematic layers into a single composite indicator through weighted summation. This method has proven effective in storing, analyzing, and visualizing information from diverse spatial layers (Faisal & Shaker, 2017). One widely adopted technique for assigning weights is the Analytic Hierarchy Process (AHP), which quantifies the relative importance of different indicators based on pairwise comparisons (Chen et al., 2010). However, the reliability of AHP heavily depends on expert judgment and prior experience, potentially leading to subjectivity (Sarkar et al., 2023). As pointed out by Nichol and Wong (2009), GIS overlay methods typically do not account for inter-parameter correlation and require manually defined class boundaries, which may reduce objectivity and accuracy. To overcome these limitations, data-driven approaches such as Principal Component Analysis (PCA) have gained attention (Cartone & Postiglione, 2021).

Among the most influential remote sensing-based approaches is the Remote Sensing-based Ecological Index (RSEI) developed by Xu (2013). This framework utilizes PCA to integrate four ecologically relevant indices: greenness (NDVI), wetness (WET), dryness (typically represented by NDBSI), and heat (LST). By taking the first principal component (PC1) as a proxy for overall ecological quality, RSEI provides an objective and consistent composite

measure of environmental conditions. Its robustness and replicability have led to wide applications across varied landscapes such as urban areas, mining zones, and river basins (Wang et al., 2023; Zhu et al., 2021).

1.2 Principal Component Analysis in Environmental Remote Sensing

Principal Component Analysis (PCA) is an unsupervised feature extraction and dimensionality reduction technique (Zabalza et al., 2014), its core principle lies in applying a linear transformation to convert a set of correlated variables into a new set of uncorrelated composite variables known as principal components (PCs). The first principal component captures the most significant variance in the data, while subsequent components, each orthogonal to the preceding ones, capture progressively less variance (Teffera et al., 2018; W. -t. Cai et al., 2010). By reducing redundancy among variables and preserving the most significant patterns of variation, PCA facilitates interpreting complex relationships between environmental indicators (Kebonye et al., 2021), supporting both analysis and visualization.

In remote sensing-based ecological assessment, PCA offers a key advantage by eliminating the need for subjective weight assignments. Instead, the relative importance of each variable is automatically determined by their loadings in the principal components, thus providing an objective and consistent method for data integration (Cartone & Postiglione, 2021; Faisal & Shaker, 2017). In recent years, the integration of PCA with various remote sensing indices has become increasingly common. For instance, Teffera et al. (2018) applied PCA in conjunction with multiple spectral indices such as NDVI, NDWI, and NDBI to enhance land cover classification accuracy in the Middle Awash sub-basin of Ethiopia. Their findings demonstrated that using composite index imagery followed by PCA processing improved classification outcomes and enabled more precise differentiation of land use types, particularly in geologically complex and rapidly changing landscapes. Similarly, Faisal and Shaker (2017) employed PCA for urban environmental quality (UEQ) evaluation, comparing it with a traditional GIS overlay method. They highlighted the capability of PCA to reduce dimensionality and identify key environmental characteristics from a complex array of urban indicators. Xu (2013) also employed PCA to construct the Remote Sensing-based Ecological Index (RSEI), which remains widely used in large-scale ecological assessment.

Despite its widespread application, PCA has inherent limitations. It assumes linear relationships among variables, which may not fully capture the complex, potentially nonlinear interactions in remote sensing data (Liao & Jiang, 2020). Moreover, PCA

components are linear combinations of variables with no guaranteed physical meaning, their ecological interpretation often requires further analysis of loadings or supplementary visualization (Demšar et al., 2013; Jolliffe, 2002). Wang et al. (2023) highlighted that the explanatory power of the first principle component (PC1) typically ranges between 60% and 90%, and over-reliance on PC1 alone may lead to the loss of crucial ecological information. Zhu et al. (2021b) indicated that the limitation of the RSEI model lies in that when the study area is an extensive range, the equalization of the weights of environmental factors will not be representative.

1.3 Cartographic Visualization of Environmental Data

Environmental data can be represented through two fundamental spatial data models: vector and raster. Each model supports distinct cartographic techniques tailored to its data structure. Vector data, composed of discrete geometries (points, lines, and polygons), is typically used to represent features such as land parcels, infrastructure, or administrative boundaries (Diamond, L., 2019). Its visualization methods include choropleth maps, proportional symbol maps, and flow maps, which are effective in presenting socio-economic or categorical variables (Golebiowska et al., 2021).

In contrast, raster data consists of a continuous matrix of grid cells, each storing a single value representing a spatial phenomenon (Williams, C., 2019), commonly used in environmental remote sensing. This model excels in depicting continuous variables such as temperature, vegetation density, or soil moisture. Raster-based visualizations commonly employ color gradients, isopleth (contour) maps, or pixel-based classifications, allowing for nuanced spatial interpretation at fine resolutions (Golebiowska et al., 2021).

While vector maps dominate traditional thematic mapping, the increasing use of satellite-derived raster indices has expanded the toolkit of cartographers working in environmental domains. Raster data visualization emphasizes the continuous nature of environmental variables and often requires different color design strategies to maintain interpretability and accessibility.

However, regardless of format, the effectiveness of map-based communication depends not only on the data model, but also on the cognitive limitations of users, the design decisions of cartographers, and the context of map use. As noted by Bazaglia Filho et al. (2013) compared traditional soil maps with digital soil representations using principal component analysis (PCA) and fuzzy k-means clustering. Their findings highlight the key advantage of digital

cartography: the ability to quantify and systematize spatial parameters, thereby reducing the subjectivity and inconsistency often present in manually interpreted traditional maps.

Furthermore, cognitive research emphasizes the limited capacity of human perception in simultaneously processing visual information. According to Lucchesi and Wikle (2017), viewers can generally comprehend no more than three dimensions of visual input at any given time without cognitive overload.

Therefore, to ensure the accuracy of information and avoid the complex transmission of information, only bivariate visualization methods are discussed in this paper.

1.3.1 Univariate Thematic Mapping Method

Univariate thematic maps depict the spatial distribution of a single variable, enabling intuitive interpretation of patterns across a geographic area. In cartographic theory, commonly used univariate mapping methods include choropleth maps, dot density maps, isopleth maps, and graduated symbol map. Choropleth maps are one of the most common types of maps used in univariate thematic maps, which represent spatial differences in numerical values through gradient variations in color shades, usually based on administrative divisions or regular shaped Mosaic bases. Point density maps are used to represent the density of a certain phenomenon or event in a specific area, such as vegetation cover density or population distribution. The scale symbol plan expresses the difference of numerical size intuitively through the symbol size. Contour maps represent data through interpolated lines, and are suitable for displaying spatial gradient characteristics of continuous variables (such as temperature and precipitation) (Golebiowska et al., 2021).

While most traditional univariate maps have been developed with vector data in mind, raster data is particularly suited for representing continuous environmental phenomena such as land surface temperature, vegetation indices, or precipitation. Each raster cell corresponds to a fixed spatial location and resolution, making it ideal for conveying fine-grained variations in environmental quality. In such cases, color gradients, either continuous or discrete (**Figure 1-1**), play a critical role in map readability and interpretability.

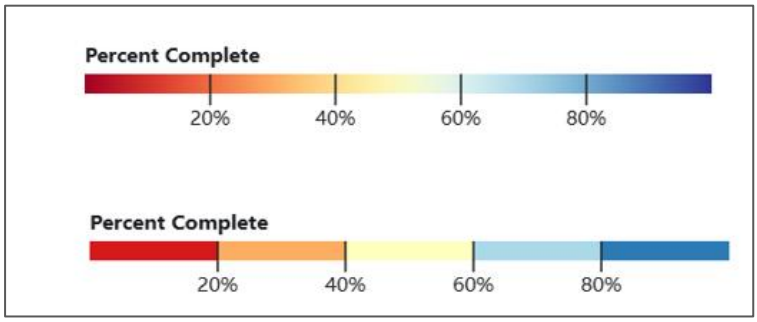


Figure 1-1 Continuous and discrete color schema (Color Scales).

Raster-based environmental maps often rely on intuitive color gradients to help non-specialists interpret complex spatial data. For example, the Global Climate Monitor platform visualizes long-term climate trends and anomalies using color schemes where red represents high temperatures and blue indicates cooler areas (Persson, 2020). Similarly, in NDVI, visualizations, green hues are commonly used to indicate vegetated regions, with darker tones implying higher vegetation density. When the goal shifts toward highlighting degraded or sparsely vegetated zones, diverging color palettes such as red-to-green schemes are often applied, where red emphasizes ecological stress (Racoviteanu et al., 2024). These examples highlight that color choices in environmental visualization are semantically meaningful and must be aligned with the specific communication objective.

Despite their simplicity and intuitive readability, univariate maps inherently lack the capacity to illustrate inter-variable relationships. For example, multiple univariate maps are often arranged as small multiples (**Figure 1-2**) to compare variable patterns (Kebonye et al., 2023). However, the limitations of univariate maps become more apparent when multiple environmental variables need to be analyzed comprehensively and when spatial correlations between them need to be explored.

Univariate maps are generally easier to interpret, but they often hinder users from engaging with spatial complexity and covariation among variables. In contrast, bivariate maps that combine two variables into a single visual output enable more effective communication of spatial interrelationships and covariation patterns, supporting more integrative interpretation (Leonowicz, 2006; Stachoň et al., 2025).

Figure 1-2 Small multiples map from Nelson and Wilson (2020).

1.3.2 Multivariable Visualization Method

Multivariable cartographic visualization techniques are designed to represent two or more variables simultaneously within the same spatial framework, enhancing the interpretation of inter-variable patterns and interactions. Among these, bivariate maps are one of the most widely used techniques, applicable to both vector and raster data formats. These maps allow the simultaneous display of two variables by merging them into a single composite visual output (Kebonye et al., 2023; Neto, 2022). Instead of showing two separate univariate maps, a bivariate choropleth employs a two-dimensional color scale (**Figure 1-3**) that reflects both variables simultaneously. As Teuling et al. (2011) explained, the bivariate composite $Z = f(X_1, X_2)$ allows for the visualization of both marginal distributions and their spatial covariance structure, which offers more nuanced insights into their interrelationships.

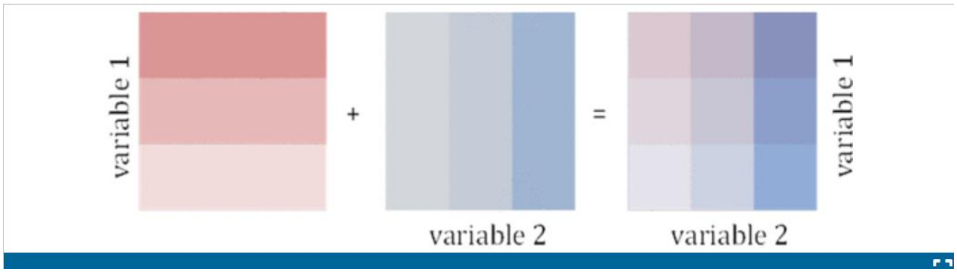


Figure 1-3 Building a bivariate color palette, from Cheng et al. (2019).

The classification of raster data for bivariate mapping typically involves categorizing each variable independently, then creating a matrix of class combinations (e.g., A1, B2, etc.). Stevens (2015) provided a widely used methodology for building bivariate color palettes and implementing them in GIS environments, including detailed procedures for QGIS. However,

the effectiveness of bivariate maps is highly dependent on color inappropriate design combinations can lead to false color perception or misinterpretation (Cheng et al., 2019).

Color is an important part of the symbol system, Wilkening et al. (2019) identified the visual attributes that attract the viewer's attention the most when it comes to the colour and movement of the map by analysing the user's eye-tracking. Persson (2020) emphasized that the choice of color representation is critical, since most people have poor color vision, the wrong color choice can lead to ineffective maps. For example, in NDVI mapping, variations within a single hue like green may be hard to distinguish. A diverging palette, such as red for low values and green for high values, can better highlight degraded areas (**Figure 1-4**).

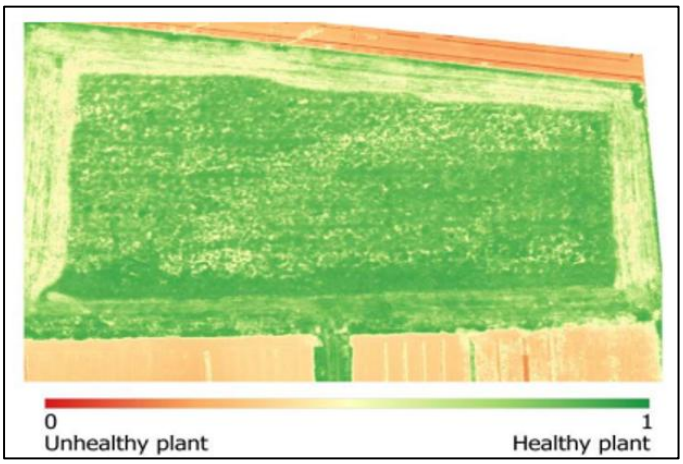


Figure 1-4 Visualization of NDVI values with a diverging color ramp, illustrating plant health conditions ranging from unhealthy (red) to healthy (green). Source: Boiarskii and Hasegawa (2019).

The common method for color model construction is to use the Red, Blue and Green (RGB) color model, by adjusting the amount of RGB or brightness and saturation, so as to create a unique color for each value (Teuling et al., 2011). Beyond hue, an important visual variable is transparency, which can be encoded using the alpha channel in image composition. Unlike CMYK, which is primarily optimized for print media and subtractive color mixing, RGB is additive and better suited for digital map displays and web-based cartographic applications (Ibraheem et al., 2012).

The Alpha channel controls pixel opacity, ranging from 0 (fully transparent) to 1 (fully opaque), and blends overlaid objects by modulating transparency in combination with RGB values. This allows fractional alpha values along object edges to produce anti-aliasing effects, while fully transparent areas are rendered black in RGB channels (Porter & Duff, 1984). This

principle is further extended in the value-by-alpha map, by adjusting the transparency, spatial units with lower values appear increasingly faded, while the higher value areas are highlighting, producing a similar "spotlight" visual guidance effect, as shown in **Figure 1-5** and described by Roth et al. (2010). The pre-multiplied Red Green Blue Alpha (RGBA) representation further improves the computational efficiency of color fusion and optimizes the visual effect and cartographic representation (Porter & Duff, 1984).

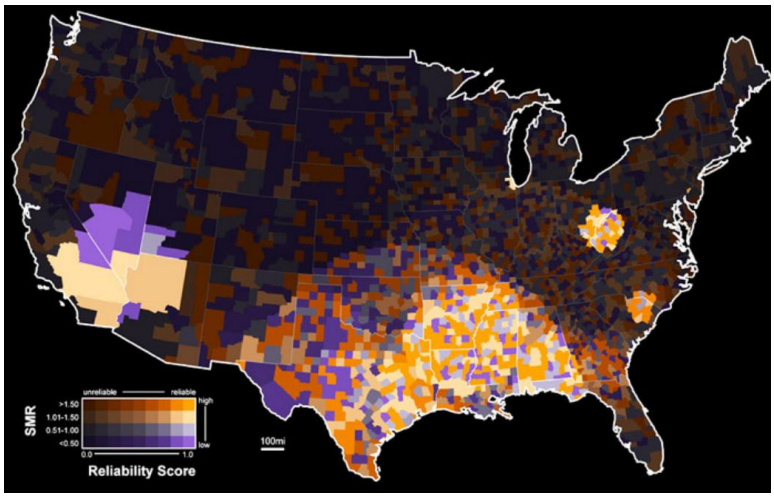


Figure 1-5 A value-by-alpha map visualizing U.S. cervical cancer mortality rates (2000–2004), where color encodes the standardized mortality ratio (SMR) and transparency reflects data reliability scores. Adapted from Roth et al. (2010).

Roth et al. (2010) proposed two alpha mapping strategies:

1. In the single-layer method, the variable of interest is mapped to RGB, and the equalizing variable is assigned to the alpha channel of the same layer. Color blending occurs between the upper layer and a fixed background, requiring careful control of the background color for perceptual clarity.
2. In the dual-layer method, the RGB values are assigned to a bottom layer, while alpha is defined in a top layer. The higher the equalizing variable, the more transparent the top layer becomes, enabling better contrast control between two layers.

Choosing the correct color (black or white) affects the visual variables after the color fusion. Black is typically preferred as the background color, as it preserves contrast better than white, which may reduce visual sharpness by simultaneously altering saturation and brightness (Roth et al., 2010).

Roth et al.(2010) provide a set of five design parameters for effective mapping of alpha values, as shown in **Figure 1-6**.

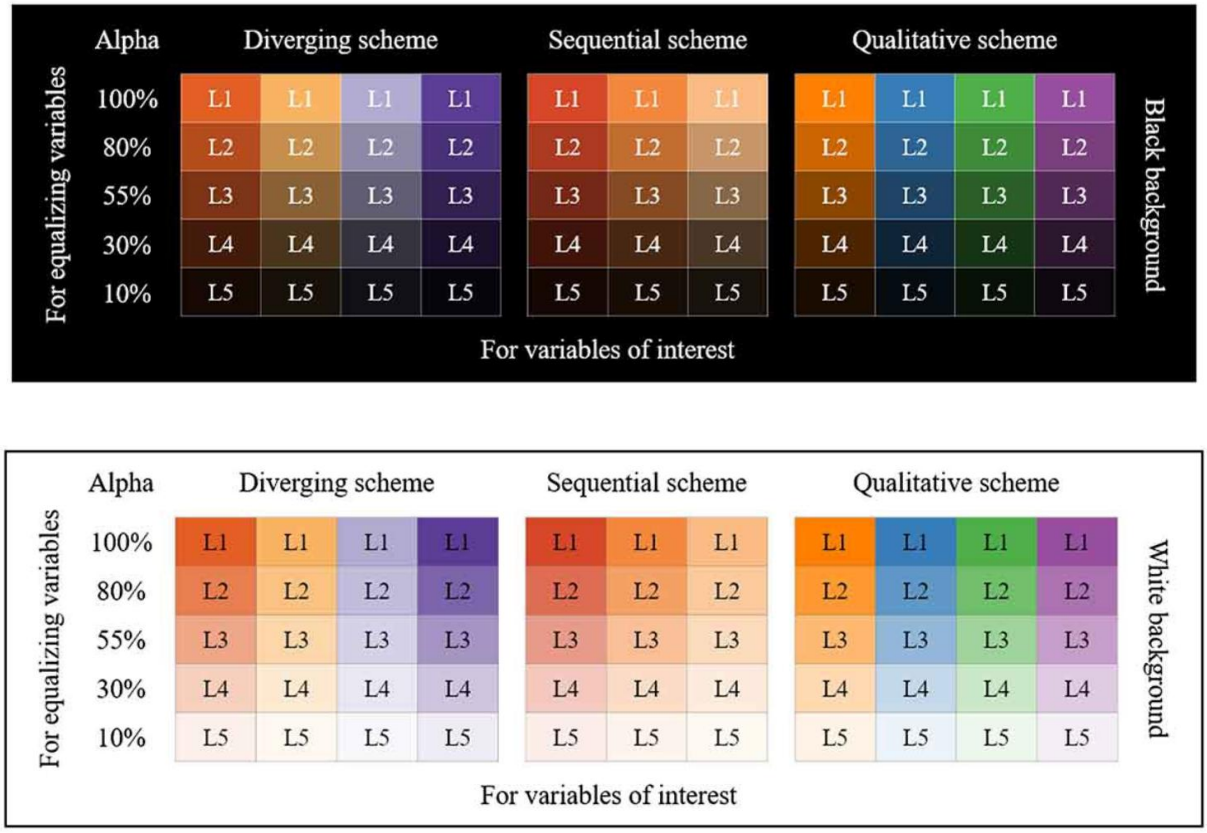


Figure 1-6 Alpha scheme for value-by-alpha mapping, black background (Top), white background (Bottom). Adapted from Roth, Woodruff, and Johnson (2010).

From the above studies, PCA has significant value in analysing remote sensing ecological indices, but its application in image expression and spatial visual communication is still relatively limited. In this study, PCA is not applied to environmental assessment but is used as a dimensionality reduction technique to support multivariate visualization of ecological indices. By transforming five correlated remote sensing-based indicators (NDVI, NDWI, NDMI, BSI, and LST) into uncorrelated principal components, PCA enables the extraction of dominant environmental gradients while reducing data redundancy. This facilitates more interpretable and compact visual representations, which are then compared across different cartographic techniques, which include weighted index maps, bivariate colour maps, and value-by-alpha visualizations to evaluate how PCA-based dimensionality reduction can enhance the effectiveness of multivariate environmental raster visualization.

2 Methodology

This study adopts a multi-step geospatial workflow that applies PCA-based dimensionality reduction to remote sensing indices and evaluates visualization strategies across four contrasting landscapes in Estonia. The methodological framework encompasses remote sensing data preprocessing, principal component extraction using Python, and the implementation of cartographic visualization techniques in QGIS and R. The entire workflow is shown in **Figure 2-1**.

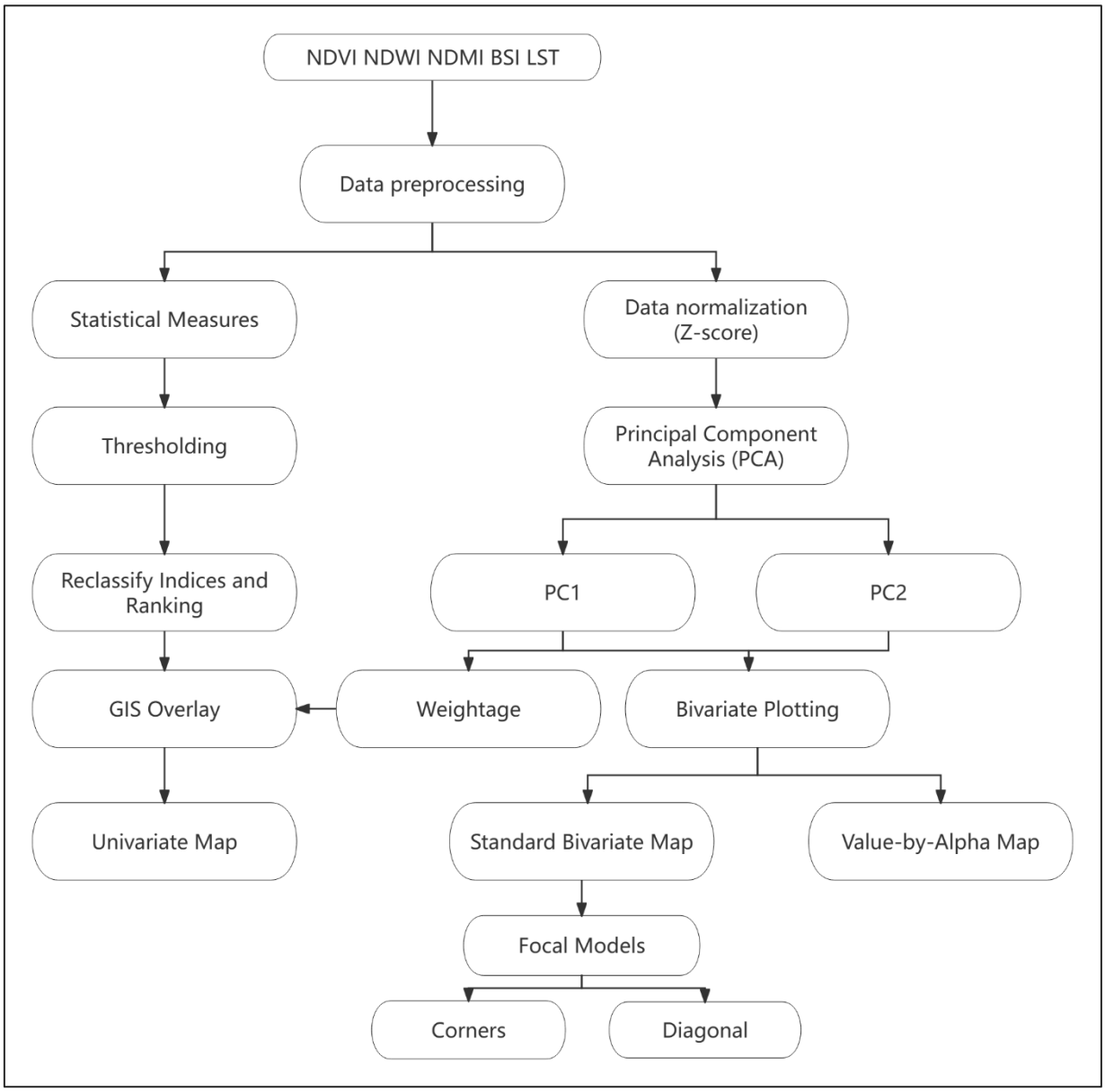


Figure 2-1 Technical circuit diagram.

The methodology consists of the following key stages:

a) **Data Preprocessing:**

A set of satellite-derived environmental indices is selected to capture major dimensions of environmental quality, including greenness (NDVI), moisture content (NDWI, NDMI), bareness (BSI), and surface temperature (LST). All indices were resampled to 10m resolution and extent and preprocessed to reflect surface conditions in the summer of 2024.

b) **Principal Component Analysis (PCA):**

To synthesize multivariate index data and extract the dominant gradients of variation, PCA was applied to standardized index layers. The analysis is conducted separately for each case study area. The first principal component (PC1) was interpreted as the primary synthetic indicator of environmental condition and was used to compute a composite sustainability index. However, the results revealed that in most regions, PC1 could only explain around 60% of the total variance, suggesting that a considerable portion of information is retained in the subsequent components (especially PC2). Therefore, to avoid oversimplification, both PC1 and PC2 are preserved for downstream visualization analysis.

c) **Composite Index Construction via GIS Overlay:**

A weighted overlay analysis is conducted in QGIS to generate a single-variable environmental sustainability map based solely on PC1. First, each index layer is reclassified using a four-level thresholding scheme based on statistical deviation from the mean. Then, the PCA-derived loading values from PC1 are used to calculate relative weights for each index, normalized to sum to one. The composite map is created by summing all reclassified layers multiplied by their corresponding weights.

d) **Bivariate Visualization in R**

To better represent multidimensional environmental conditions, two types of bivariate maps are created using both PC1 and PC2:

1. Bivariate choropleth map, where two variables are mapped using a dual-color scheme.
2. Value-by-alpha map, where PC1 is mapped through a diverging color palette, while PC2 is mapped to the transparency (alpha) channel.

This combination strategy can compare traditional superposition-based synthesis with the more advanced PCA enabled multivariable visualization, highlighting how much information can be lost when reducing environmental complexity to a single axis. Adding PC2 to a bivariate map can significantly improve interpretability, as PC1 alone is not sufficient to reflect subtle environmental dynamics. In selecting the indices for analysis (NDVI, NDWI, NDMI, BSI, and LST), both theoretical relevance to environmental conditions and practical considerations such as data availability and temporal consistency were considered. In contrast to earlier frameworks such as Xu's (2013) RSEI or urban-focused studies like Sharma and Joshi (2016) , which utilize the Normalized Difference Built-up Index (NDBI), this study replaces NDBI with the Bare Soil Index (BSI). This decision was informed by two main factors: first, the lack of temporal alignment between available NDBI data and other indices, which could compromise the coherence of PCA input; and second, a conceptual shift away from purely urban emphasis. As our study spans a gradient of urban and non-urban landscapes, including forested catchments and post-mining environments, BSI provides a more suitable proxy for surface dryness and land exposure across heterogeneous land cover types, extending beyond the built-up surfaces captured by NDBI.

2.1 Study Area

This study focuses on four representative regions across Estonia, selected to capture a diversity of landscape types and environmental conditions: t the densely built-up coastal capital Tallinn, the lower-density inland city of Tartu, the ecologically varied Porijõgi catchment in southeastern Estonia, and the industrially impacted Ida-Viru Mining Region in the northeast. These areas are not confined to administrative boundaries but are delineated to reflect dominant landscape characteristics and ecological heterogeneity. As illustrated in **Figure 2-2**, these sites span distinct geographic zones and landscape types, providing a varied spatial context for environmental visualization.

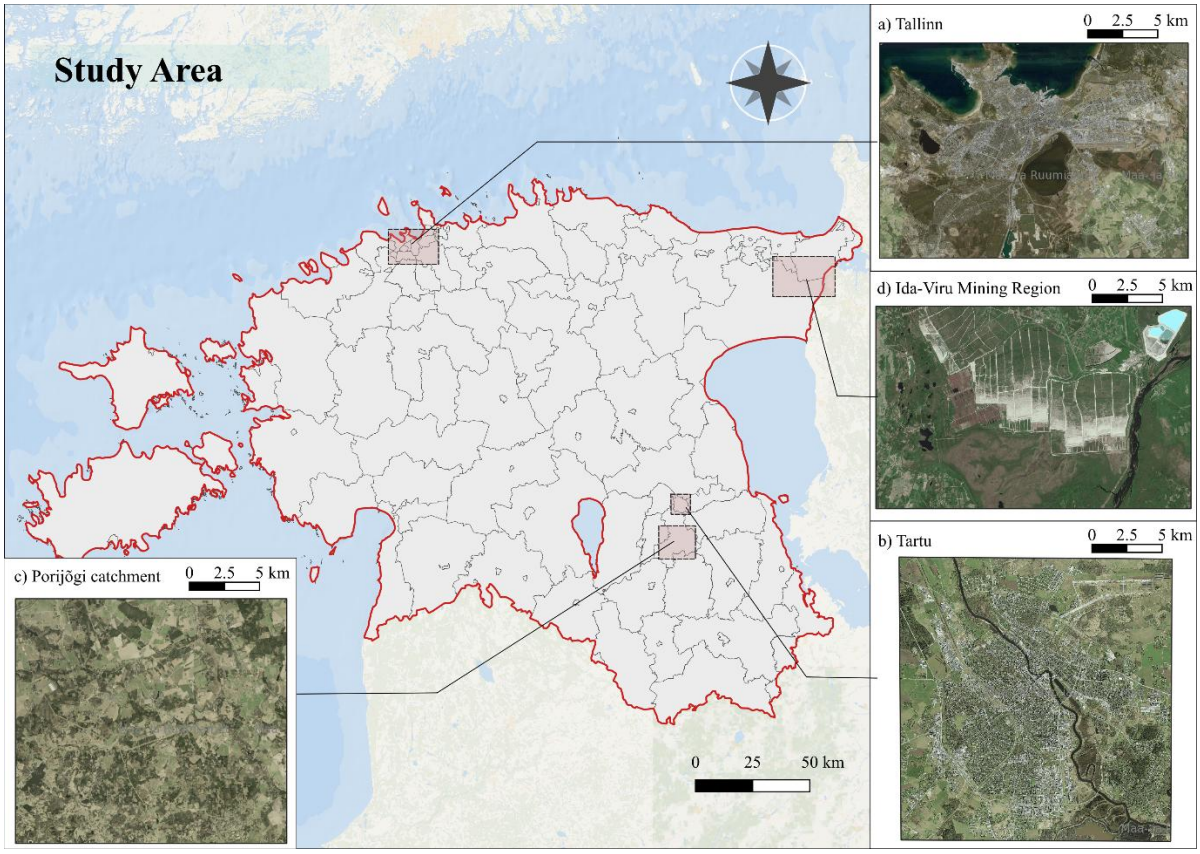


Figure 2-2 Study Area map displays the four selected study areas across Estonia, each representing a distinct landscape type. (a) Tallinn; (b) Tartu; c) Porijõgi Catchment; (d) Ida-Viru Mining Region. The main map outlines the national territory and administrative boundaries of Estonia, with red lines indicating the national border and pink shaded areas marking the locations of the four case study sites.

2.1.1 Tallinn (High-density Urban Landscape)

Tallinn, the capital of Estonia, is situated along the northern Baltic Sea coast (59°26′N, 24°45′ E). The city is characterized by dense urban development and coastal ecosystems, with flat terrain and the highest elevation of only 63 meters. Lake Ülemiste is the largest water body within the area. Tallinn experiences a humid continental climate, featuring warm to mildly hot summers and cold, snowy winters (Eslamirad et al., 2023). According to the Estonian Environment Agency (2020), the average annual precipitation at the Tallinn-Harku station between 1991 and 2020 was approximately 700 mm. Tallinn has a humid continental climate, with cold, snowy winters and mild to warm summers. The average temperature in January is around -2,9°C, while in July, it is 17.6°C (Estonian Environment Agency, 2020). Urban land cover dominates the landscape, with impervious surfaces accounting for around 50% of the total area (Laroche, 2019; Maharjan et al., 2016).

2.1.2 Tartu (Lower-density Urban Landscape)

Tartu, the second-largest city in Estonia, is located in the southeastern part of the country along the Suur-Emajõgi River (58°22′N, 26°43′E). The urban landscape integrates extensive green spaces, including several large parks, meadows, and forested areas. Long-term climatic records show that the average temperature in January is approximately -6.6°C and 17.1°C in July (Eensaar, 2016). The region receives about 673 mm of annual precipitation (Estonian Environment Agency, 2020), contributing to a relatively stable hydrological and ecological regime.

2.1.3 Porijõgi Catchment (Diverse Watershed Landscape)

The Porijõgi catchment is in southeastern Estonia, center approximately at 58.22°N latitude and 26.69°E longitude. The catchment lies within Tartu County and forms part of the larger Emajõgi River basin, contributing to the hydrological system draining into Lake Peipsi (Moges et al., 2022). The area is topographically fragmented, comprising the town of Kambja, the Palumägede matkarada nature trail, and a mosaic of small villages and forest patches. Climatically, the area experiences cold winters and mild summers, with average temperatures reaching -7.92 °C in January and approximately 22.5 °C in July. The mean annual precipitation is around 600mm (Estonian Environment Agency, 2020).

2.1.4 Ida-Viru Mining Region (Industrial Landscape)

The northeastern Ida-Viru Mining Region is near Ida-Virumaa County and represents a landscape heavily influenced by industrial activities. It includes the nature conservation area (rich in forested lakes), the Viivikonna mining complex, and the Ida-Virumaa settebasseinid, a sedimentation basin formed by the operational activities of local power plants. The central coordinate is approximately 59.24°N, 27.75°E, with a maximum elevation of 82 meters. Based on climate records from the nearby Jõhvi city, the average January temperature is around -4,6 °C, while the July average reaches 17.4°C. The area receives about 717 mm of precipitation annually (Estonian Environment Agency, 2020)

2.2 Data

This study utilized five remote sensing-derived indices to capture various environmental dimensions: greenness (NDVI), canopy moisture (NDMI), surface moisture (NDWI), surface bareness and imperviousness (BSI), and surface temperature (LST). These indices are used not as standalone ecological indicators but as test variables to demonstrate the effectiveness of various visualization techniques explored in this study.

2.2.1 Normalized Difference Vegetation Index (NDVI)

The Normalized Difference Vegetation Index (NDVI) is one of the earliest and most widely used remote sensing-derived vegetation indices. It has become a standard product for vegetation assessment. NDVI is based on the normalized reflectance difference between near-infrared (NIR) and red (RED) spectral bands, capturing the distinct reflectance characteristics of healthy vegetation in these wavelengths(Patil et al., 2024; Xue & Su, 2017).

NDVI is commonly used to monitor the growth status or activity of the vegetation canopy and, by extension, to assess surface vegetation conditions. The distribution and magnitude of NDVI values can effectively describe vegetation cover and density, playing a critical role in ecological balance monitoring.

It is mathematically defined as:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Here, NIR and RED denote the reflectance values in the near-infrared and red portions of the electromagnetic spectrum, respectively(Chandrasekar & Sesha Sai, 2015; Patil et al., 2024).

When using Landsat 8 imagery, NDVI is calculated as:

$$\frac{(Band\ 5\ -\ Band\ 4)}{(Band\ 5 +\ Band\ 4)}$$

NDVI exhibits values within a range of -1 to 1 and serves as a critical indicator of vegetation coverage and vigor. High NDVI values approaching 1 typically indicate dense vegetation coverage, positively correlated with soil moisture retention capacities. Conversely, values approaching -1 signify non-vegetated surfaces, such as urban areas, open terrains, and water bodies (Artikanur et al., 2022).

Annatakarn et al (2022) conducted a comprehensive analysis and reclassification of NDVI thresholds in Thailand, establishing distinct NDVI intervals. Their categorization is as follows:

- Negative values (approximately -1): indicative of water bodies.
- Values close to zero (-0.1 to 0.1): associated predominantly with barren or unvegetated surfaces.
- Low positive values (approximately 0.2 to 0.4): reflective of shrublands and grassland.

- High positive values (approaching 1): indicative of dense temperate and tropical rainforest ecosystems, serving effectively as proxies for healthy, green vegetation (Annatakarn et al., 2022).

In a related study, Boiarskii and Hasegawa (2019) investigated NDVI imagery from the experimental farm at Niigata University, Japan. Their analysis revealed distinct green spectral signals across vegetated areas but notably weaker signals within crop areas. The researchers categorized NDVI into two primary intervals to assess vegetation health:

- Lower to mid-range values (0 to intermediate values): indicative of unhealthy or stressed vegetation.
- Higher values (approaching 1): indicative of healthy vegetation.

The comparative findings from these studies emphasize NDVI's effectiveness as an indicator of vegetation condition, while also highlighting variations in threshold categorizations driven by regional ecological contexts and specific research objectives.

2.2.2 Normalized Difference Water Index (NDWI)

The Normalized Difference Water Index (NDWI) was initially developed to monitor changes in surface water content. Given that water strongly absorbs light across the visible to shortwave infrared (SWIR) regions of the electromagnetic spectrum, Ji et al. (2009) demonstrated that NDWI computed using the green and SWIR bands yields robust threshold stability. The formula expresses as:

$$\frac{(NIR - SWIR)}{(NIR + SWIR)}$$

When applied to Landsat 8, which provides two SWIR bands, the recommended computation for NDWI uses Band 5 (NIR) and Band 7 (SWIR2), as noted by Patil et al. (2024):

$$\frac{(Band\ 5 - Band\ 7)}{(Band\ 5 + Band\ 7)}$$

NDWI ranges from -1 to 1 and is widely used to identify and distinguish water bodies from terrestrial surfaces. Generally, positive NDWI values represent water features, while negative or zero values correspond to soil and terrestrial vegetation (EOS DATA ANALYTICS, 2021). Sentinel Hub (2024) notes that NDWI values greater than 0.5 typically correspond to water bodies, whereas vegetation often shows significantly lower values, and built-up areas yield values between zero and 0.

Although NDWI values greater than 0 theoretically indicate water presence, practical considerations suggest variations due to environmental factors. For instance, Zheng et al. (2021) identified sediment-laden water bodies, such as rivers and lakes, which may yield negative NDWI values, leading them to recommend a threshold of -0.1 for practical water body identification. Furthermore, McFeeters (2013) highlighted limitations in urban environments, where NDWI positives may not exclusively represent water features but also include urban surfaces like car parks with standing water, flood basins, and certain residential roofs. Additionally, shading effects can produce low NDWI values even on actual water surfaces, emphasizing the necessity for site-specific threshold calibration in urban contexts.

2.2.3 Bare Soil Index (BSI)

Bare soil index was used to identify areas of bare soil by using the spectral differences between soil and vegetation. The higher the BSI value, the area where the soil is exposed. The formula is (Panahi et al., 2024):

$$\frac{(\mathrm{SWIR2} + \mathrm{R}) - (NIR + B)}{(\mathrm{SWIR2} + \mathrm{R}) + (NIR + B)}$$

SWIR2 (short-wave infrared 2), Red (red light), NIR (near infrared) and Blue (blue light) represent the corresponding bands of satellite images respectively.

BSI is used in land cover classification, soil erosion monitoring, desertification assessment and other fields. For example, in agricultural research, BSI can be used to identify areas of bare soil in cultivated land, helping to assess the impact of soil quality and management practices. In addition, BSI can assist in the detection of new exposed surface areas during urban expansion, providing support for urban planning. Although BSI has advantages in identifying bare soil, there may be difficulties in distinguishing bare soil from certain features with similar spectral characteristics (e.g. buildings, dry vegetation remains). Moreover, factors such as atmospheric conditions, soil moisture, and organic matter content may affect its accuracy. Therefore, in practice, additional or improved indices (e.g. NDVI, NDBI, or MBI) are often used to enhance classification precision (Mzid et al., 2021; Nguyen et al., 2021).

In a recent study by Panahi et al. (2024), BSI was found to overestimate bare soil coverage, with several built-up areas misclassified as exposed soil. While this can be a limitation in some cases, such a tendency proves beneficial for our study, which focuses on mapping heterogeneous urban and peri-urban environments.

2.2.4 Normalized Difference Moisture Index (NDMI)

The Normalized Difference Moisture Index (NDMI) is a vegetation index highly correlated with canopy water content and has been demonstrated to outperform NDVI in tracking variations in vegetation biomass and water stress (Jin & Sader, 2005). It is computed using the difference and sum of reflectance values in the near-infrared (NIR) and short-wave infrared (SWIR1) regions, following the formula (Taloor et al., 2021):

$$\frac{(NIR - SWIR1)}{(NIR + SWIR1)}$$

Malakhov and Tsychuyeva (2020) reviewed the conceptual origins, historical development, and applications of the NDMI. The index was originally introduced by Gao (1996) under the name Normalized Difference Water Index (NDWI) to estimate vegetation water content (VWC). To distinguish it from other NDWI formulations, particularly that of McFeeters (1996) for open water detection, Wilson and Sader (2002) proposed renaming Gao's index to NDMI. Subsequently, Yilmaz et al. (2008) emphasized the importance of consistent terminology, advocating for the use of NDMI when referring to indices derived from NIR and SWIR bands for vegetation moisture monitoring.

As summarized by Malakhov and Tsychuyeva (2020), NDMI and its variants have been widely and successfully applied in vegetation classification, drought monitoring, and ecological condition assessment, particularly due to their sensitivity to vegetation water content and stress responses.

2.2.5 Land Surface Temperature (LST)

Land Surface Temperature (LST) is a key variable influencing long-wave radiation and turbulent heat flux at the land–atmosphere interface, playing a critical role in the surface energy and water balance from local to global scales (Z.-L. Li et al., 2013). I Land Surface Temperature (LST) is a key variable influencing long-wave radiation and turbulent heat flux at the land–atmosphere interface, playing a critical role in the surface energy and water balance from local to global scales (An et al., 2022). In this study, LST was derived from the Landsat Level-2 Surface Temperature product provided by the United States Geological Survey (EROS Center, 2020). This Collection 2 dataset is atmospherically corrected and integrates TOA reflectance, brightness temperature, NDVI-based emissivity adjustments, and meteorological profiles such as geopotential height and humidity. The product meets CEOS

Analysis Ready Data (ARD) standards, ensuring consistency and compatibility for large-scale analysis.

For the LST, raw digital number (DN) values were converted to physical temperature in Celsius degrees. This conversion followed the official metadata specification provided with the dataset. Specifically, the DN values were first transformed into Kelvin using the formula:

$$T(K) = DN \times Scale\ Factor + Additive\ Offest$$

where the scale factor was 0.00341802 and the additive offset was 149. The resulting Kelvin temperatures were subsequently converted to degrees Celsius using the relation:

$$T(℃) = T(K) - 273.1$$

2.2.6 Data Sources

The indices NDVI, NDWI, NDMI, BSI were acquired from the Landscape Geoinformatics Lab, primarily derived from Sentinel-2 and Landsat 8 imagery.

The LST datasets were obtained from the USGS Earth Explorer platform, based on Landsat 8-9 (OLI/TIRS) Collection 2 (C2) Level 2 Science Product (EROS Center, 2020). For all LST indices, images were selected to represent the summer season (June to August 2024) with minimal cloud cover, to ensure consistent and reliable observations across the study areas. The detailed information of data is shown in **Table 2-1** Data.

Table 2-1 Data Source.

ID	Index	Date of acquisition	Resolution	Source
est_s2_ndvi_2024	NDVI	2024-06-01_2024-08-31	10m	Sentinel-2
est_s2_ndwi_2024	NDWI	2024-06-01_2024-08-31	10m	Sentinel-2
est_s2_ndwi_2024	BSI	2024-06-01_2024-08-31	10m	Sentinel-2
est_s2_ndmi_2024	NDMI	2024-06-01_2024-08-31	10m	Sentinel-2
LC08_L2SP_187019_20240627_20240709_02_T1_ST_B10	LST	2024-06-27_2024-07-09	30m	Landsat-8
LC09_L2SP_186019_20240628_20240702_02_T1_ST_B10	LST	2024-06-28_2024-07-02	30m	Landsat-9

2.2.7 Data Preprocessing

All raster layers (NDVI, NDWI, NDMI, BSI, and LST) were reprojected to the Estonian national CRS (EPSG:3301) and clipped to the Tallinn study area using administrative

boundaries. To ensure spatial consistency, all raster data were resampled to 10m resolution and aligned. Invalid values (e.g., NoData, NaN, or infinite) were masked out, and the remaining pixel-wise values were standardized using z-score normalization. The processed indices were stacked and used for principal component analysis (PCA) to reduce dimensionality and highlight dominant patterns across environmental variables.

2.3 Principal Component Analysis for Multivariate Environmental Visualization

Five standardized indices (NDVI, NDWI, NDBI, NDMI, and LST) were subjected to PCA to reduce the dimensions and extract the most informative composite gradients. The analysis was implemented using the Python library scikit-learn, and the first two principal components (PC1 and PC2) were retained for further visualization as they can typically explain more than 80%-90% of the total variance. The first principal component (PC1) is interpreted as a comprehensive environmental condition index, reflecting the maximum proportion of the total variance (generally accounting for more than 50% of the total variance). The code for PCA follows the implementation adapted from ESRI Python API (Elif Bulut & Helen Thompson, 2023).

Due to the non-uniqueness of eigenvector directions in PCA and their impact on the spatial distribution of results, it was necessary to manually adjust the directions based on the sign of key environmental indicators (Ning et al., 2020).

To avoid confusion and ensure interpretability, we adopted a standard practice commonly used in ecological assessment, where PCA-derived indices are adjusted either by applying 1–PC or by multiplying PC by –1. This adjustment ensures that higher values consistently represent more favorable ecological conditions.

In this study, multiplying PCs by -1 when the principal component loading showed a negative correlations with ecologically beneficial indicators, NDVI, we multiplied PCs by -1to align the directionality, as higher vegetation coverage is generally associated with better ecological quality (Liu et al., 2022). Subsequently, the adjusted PCs were normalized to a [0,1] range using min-max normalization:

$$PCs_{normalized} = \frac{PCs - PCs_{min}}{PCs_{max} - PCs_{min}}$$

Where PCs_{max} and PCs_{min} represent the maximum and minimum values in PC across all study areas.

2.4 Visualization Method Comparison

Based on the results of PCA, three different visualization techniques are compared to highlight the spatial variations of environmental indicators. The goal is to evaluate how effectively each method conveys multivariate information from remote sensing data.

2.4.1 GIS-Based Weighted Overlay

To produce a single-variable environmental sustainability index, a weighted overlay analysis was conducted using QGIS, following the standard PCA-based integration method proposed in previous geospatial studies (Sathyaseelan et al., 2023). This approach combines multiple reclassified raster indicators into a single composite layer, enabling spatial evaluation of environmental conditions across heterogeneous landscapes.

First, the input indices (NDVI, NDWI, NDMI, BSI, and LST) were reclassified into four ordinal categories (Very High, High, Moderate, and Low) using the mean and standard deviation of each raster. The classification thresholds were defined as follows:

- **Very High (I):** values greater than Mean + 1.5 × SD
- **High (II):** Mean + 0.5 × SD to Mean + 1.5 × SD
- **Moderate (III):** Mean – 0.5 × SD to Mean + 0.5 × SD
- **Low (IV):** values less than Mean – 0.5 × SD

These thresholds are selected to ensure a balanced representation of data distribution and variability across the study area.

Weights for each index are calculated based on the loading values of the first principal component (PC1). The use of PCA allowed objective determination of index contributions, where the weight wi for index i was computed as:

$$\boldsymbol{wi} = \frac{loadingi}{\sum loadings}$$

In cases where NDVI showed negative correlation to PC1, the sign of PC1 was reversed to ensure consistency in interpretation (i.e., higher values indicating more favorable environmental conditions). All reclassified layers were then multiplied by their respective normalized weights using the raster calculator function in QGIS. The final composite map was derived by summing all weighted layers, producing a single-value raster that spatially represents relative environmental sustainability across the study regions.

2.4.2 Bivariate and Alpha Mapping

To explore the spatial co-variation of principal components (PC1 and PC2), this study applies both bivariate choropleth mapping and value-by-alpha mapping techniques. The visualization design draws upon established typologies and experimental comparisons of color logic to support interpretability and emphasize spatial patterns.

- **Bivariate Mapping Design**

To guide the visual emphasis in bivariate choropleth mapping, this study adopts the focal model framework proposed by Strode et al. (2019), which operationalizes Trumbo's (1981) original principles for effective bivariate map design. The focal models: Corners, Range, and Diagonal, each tailored to emphasize different types of spatial relationships. Inverse relationships (Corners), the distribution of one variable within a category of another (Range), and direct relationships (Diagonal). These models help align map design with map purpose, focusing attention on variable interactions via appropriate color organization and spatial emphasis (**Figure 2-3**). These models guide how two variables (PC1 and PC2) are jointly represented in a two-dimensional color grid (typically 3×3 or 4×4).

In this study, the Corners and Diagonal models were implemented using color palettes developed in line with Strode et al. (2019) via the Bivariate Color Generator tool by Angie Hjort on ObservableHQ. (https://observablehq.com/@angiehjort/bivariate-color-generator), While the Range model was not adopted due to its known limitations in balancing order and separation, value-by-alpha mapping (Roth et al., 2010) was used instead to highlight magnitude (PC1) and contextual variation (PC2) through color and transparency blending. As noted by Strode et al., value-by-alpha can outperform range-based palettes by providing clearer visual emphasis without compromising perceptual differentiation.

Although a fully customized palette tailored to the environmental context of the case study regions is not yet developed, future work may involve designing such schemes independently.

In the current stage, a standard bivariate choropleth map is also produced using the R package biscale, serving as a baseline for comparison with the focal model-based visualization. This comparison helps evaluate the effectiveness of the focal model in emphasizing specific data patterns.

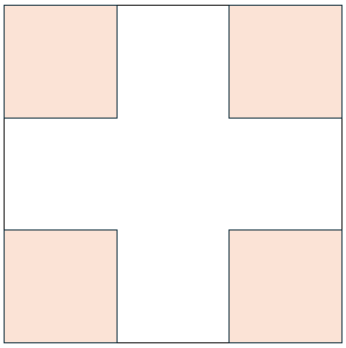
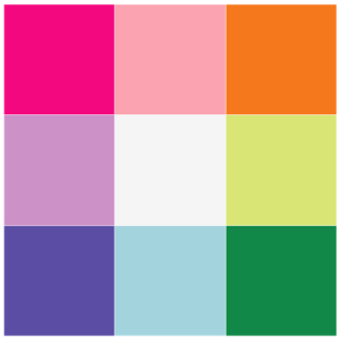
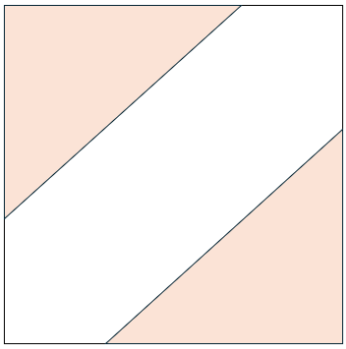
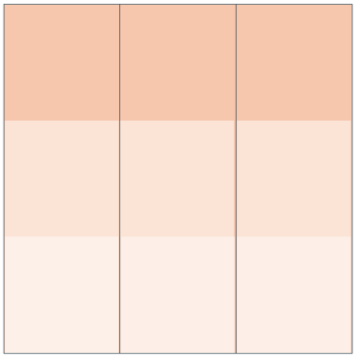
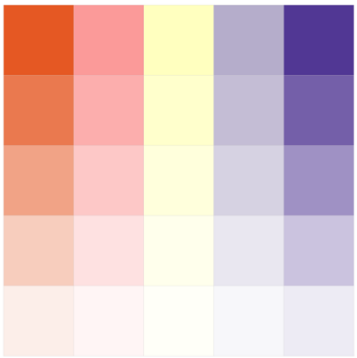
Model Name	Focal areas	color scheme
Corners		
Diagonal		
Range		

Figure 2-3 Charts adapted from Strode et al.(2019), Operationalizing Trumbo's Principles of Bivariate Choropleth Map Design.

- **Classification Methods**

Both quantile and equal interval classification strategies were considered for assigning class values to PC1 and PC2:

- Quantile classification helps distribute pixels evenly across the map, highlighting contrast across regions.
- Equal interval classification supports more intuitive thresholding, especially in less diverse areas.

The choice of classification method and color model will be evaluated through experimental mapping and visual comparison in the results section.

- **Value-by-alpha Mapping**

Mapping value-by-alpha map is essentially a bivariate choropleth technique that uses transparency (alpha channel) to encode secondary variables, allowing the primary variable to dominate color perception while still including secondary information. In this study, we adopted the diverging color scheme with alpha transparency recommended by Roth et al. (2010) . While the conceptual framework and alpha steps (100%, 80%, 55%, 30%, and 10%) follow Roth's model, the final color palette was adjusted to ensure optimal visual contrast and perceptual clarity over varied land surface backgrounds. The resulting palette retains the structural logic of Roth's approach but incorporates a customized hue set to better fit the remote sensing-based environmental gradients explored in this study.

We explored two rendering methods:

1. Alpha mapping with only transparency and color changes: PC1 is represented using a single hue sequence or divergent color gradients (such as red or blue), while PC2 adjusts the opacity from completely transparent (low value) to completely opaque (high value).
2. Alpha mapping with spotlight effect: The color is still determined by PC1, but the transparency is rendered as a mixture of black and the base color, following Roth's mixed alpha strategy. This can make the low PC2 regions appear faded or "unimportant", thereby visually prioritizing the high alpha regions.

The final color selection will be determined after experimental trials. The aim is to balance the aesthetic quality and legibility of patterns across regions with differing data distributions. By comparing these visual methods, the study seeks to identify the most effective visualization strategy to communicate multidimensional environmental patterns at the regional scale.

3 Results

3.1 PCA Results

PCA results across the four study areas indicate notable regional variation in the dimensional structure of environmental data. In Tallinn and the Ida-Viru Mining Region, both PC1 and PC2 have eigenvalues above 1 and explain over 85% of the total variance combined, suggesting the influence of two relatively independent environmental gradients. In contrast, Tartu and the Porijõgi catchment show a more dominant first component, with PC1 alone explaining the majority of variance, indicating more cohesive environmental patterns. These structural differences imply varying levels of ecological complexity and heterogeneity across urban, diverse watershed and industrial landscapes. Detailed numerical results are presented in Appendix **Table A-1** to

Table A-4 and the corresponding visualizations of PC1 and PC2 distributions for each region are shown in **Figure A-22** to **Figure A-25**.

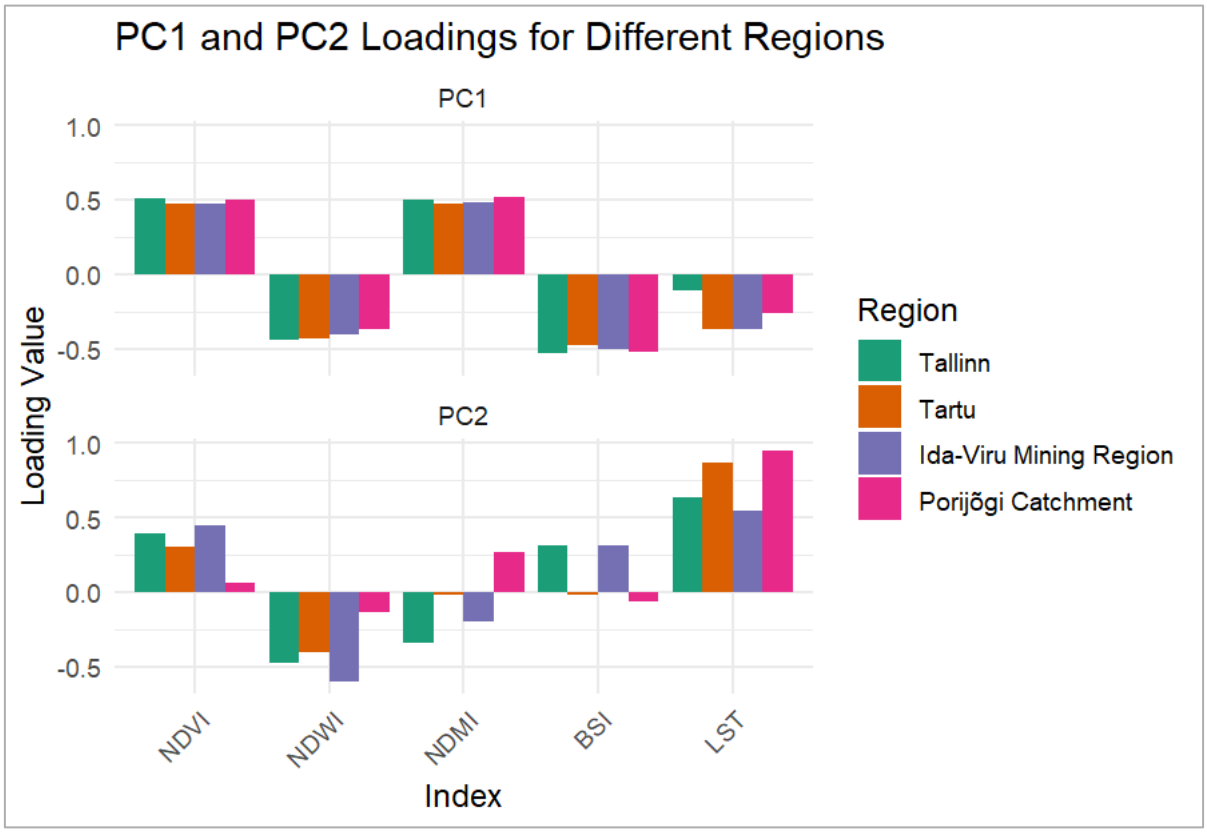


Figure 3-1 Loadings of five remote sensing indices on PC1 and PC2 across the four study regions.

Figure 3-1 presents the loadings of five remote sensing-based environmental indices (NDVI, NDWI, NDMI, BSI, and LST) on the first two principal components (PC1 and PC2) across the four study areas. These loadings reflect the relative contribution of each index to the underlying environmental variation and support the interpretation of the dominant environmental gradients identified through PCA.

In all regions, PC1 distinguishes vegetated and moisture-rich environments from bare or anthropogenically disturbed areas. NDVI and NDMI show strong positive loadings on PC1, reflecting their association with vegetation density and moisture. In contrast, NDWI and BSI contribute negatively, indicating bare soil, built-up surfaces, or water body. This component can be interpreted as a vegetation–degradation gradient.

By contrast the load structure of PC2 shows more significant differences among different regions. For instance, LST shows stronger positive loading in Tartu and Porijõgi catchment, lower in Tallinn and Ida-Viru Mining Region. The load direction and intensity of PC2 in NDWI and NDMI also vary in different regions. For instance, in Tallinn and Tartu, the loads of NDWI and NDMI are negative, while the situation is different in the Porijogi catchment. BSI shows positive load on PC2 in the Ida-Viru Mining Region and the Porijogi catchment, while it shows negative or nearly zero load in Talline and Tartu. These differences indicate that PC2 captures region-specific variation patterns, which may be related to secondary factors such as unique surface cover types, hydrological conditions or thermal environments in various regions. These secondary factors have different combined effects on the index in different regions.

Notably, NDWI consistently loads negatively on both PC1 and PC2, underscoring its stable role in identifying liquid water across different environmental gradients. These patterns indicate that PCA effectively isolates key environmental structure vegetation condition and thermal stress to support its utility for ecological assessment. For detailed eigenvalues and the proportion of variance explained by each principal component, refer to Appendix

Table A-5 to

Table A-8.

3.2 GIS Overlap Results

This section presents the environmental sustainability zones generated through a GIS-based weighted overlay method, utilizing five remote sensing indices: NDVI, NDWI, NDMI, BSI, and LST. The classification thresholds and weight distributions for these indices were determined based on regional environmental characteristics and varied across the four study areas (Tallinn, Tartu, Porijõgi catchment, and Ida-Viru Mining Region). Detailed classification and weighting results for each index and region are provided in Appendix (

Table A-9 to

Table A-12).

Figure 3-2 presents the Environmental Sustainability Zones in Tartu for the year 2024. The spatial distribution clearly delineates areas of varying sustainability levels. Very High sustainability zones (dark green) are predominantly located along the river corridor, in peripheral forested or vegetated areas, and in parklands. These areas are characterized by high greenness, surface moisture, and lower levels of bareness and land surface temperature, reflecting robust ecological conditions. This spatial pattern is consistent with the strong positive loadings of NDVI and NDMI on PC1, and negative loadings of BSI and LST on PC1, as shown in **Figure 3-1**, indicating that the GIS overlay map effectively captures the dominant environmental gradient represented by PC1. In contrast, Low sustainability zones (red) are visibly concentrated within the central built-up areas of Tartu, particularly in districts exhibiting high BSI (bare soil index) and LST (land surface temperature) values. Moderate sustainability zones (yellow) typically represent transitional areas between the urban core and natural green spaces, or regions with mixed land uses.

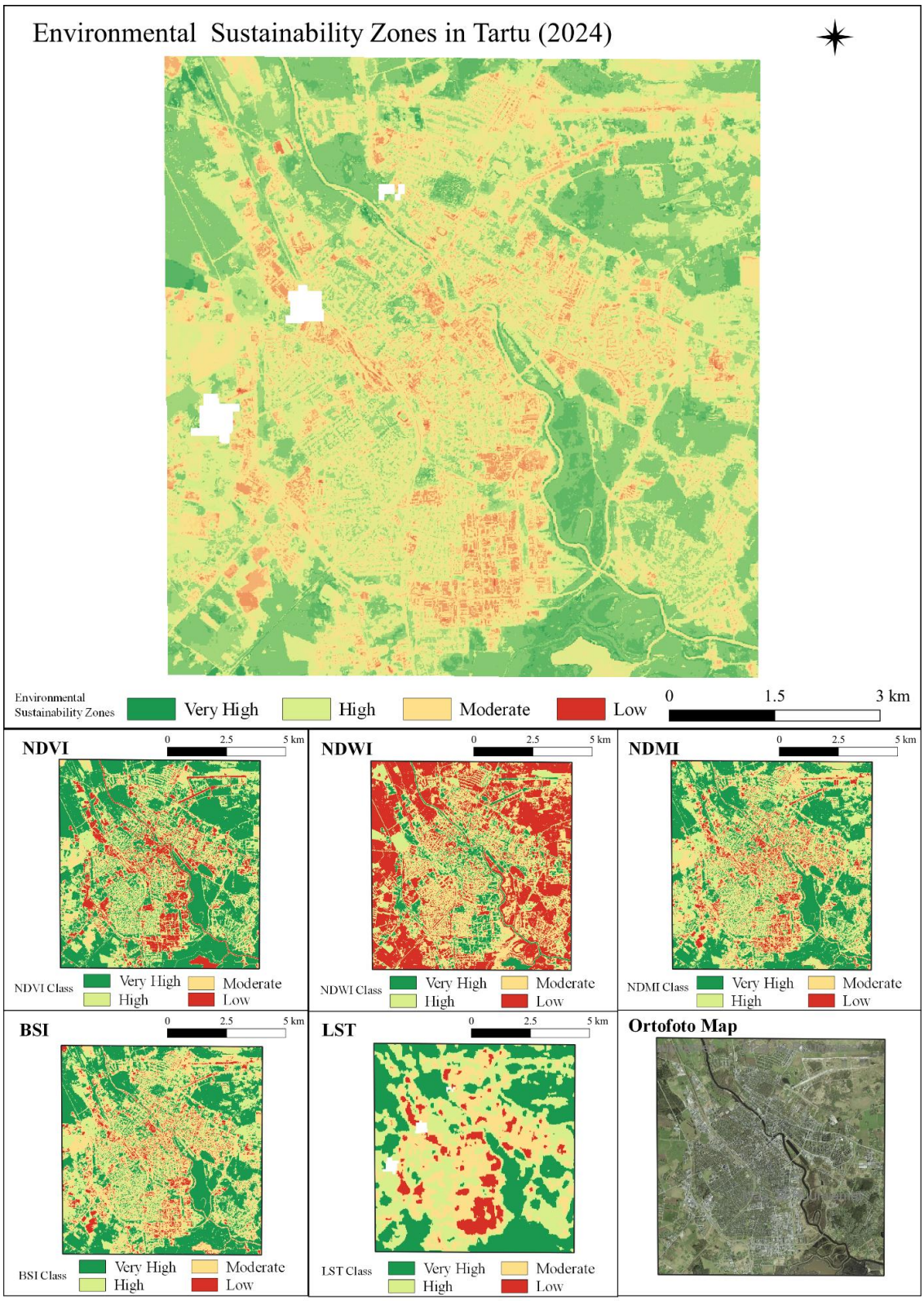


Figure 3-2 The environmental sustainability zones map in Tartu (2024). The main map presents the composite environmental sustainability zones in the Tartu region, derived from five ecological indices: NDVI, NDWI, NDMI, BSI, and LST. The lower panel shows the individual classified maps of each input index and orthophoto map.

3.3 Bivariate Raster Visualization Results

To facilitate interpretation of the PCA-based bivariate maps, **Table 3-1** categorizes the ecological meaning of the PC1 and PC2 combinations across the four study areas. Each quadrant reflects distinct environmental conditions derived from the dominant PCA axes, providing a clearer spatial understanding of urban, vegetated, and moisture-related gradients in the mapped landscapes.

Table 3-1 Environmental interpretation of PC1 and PC2 combinations in four study areas

Region	**PC1 High / PC2 High**	**PC1 High / PC2 Low**
Tallinn	Vegetated urban zones with warm microclimates	Dense vegetation with high moisture
Tartu	Urban green spaces with elevated temperatures	Forested or irrigated vegetated zones
Porijõgi	Plant cover but warm, moist zones (open vegetated)	Plant cover and low temp (forested or wetland)
Ida-Viru Mining Region	Plant cover but warm and low moist (regeneratively land)	Moist green areas, water-edge zones
Region	**PC1 Low / PC2 High**	**PC1 Low / PC2 Low**
Tallinn	Urban or built-up with thermal stress	Barren heated lands (e.g., bare mining surfaces)
Tartu	Urban core or dry lawn	Water bodies, bareness surface but low temperature
Porijõgi	Exposed sunlit areas, bare ground and high temp	Bare ground, low temp (shaded moist zones)
Ida-Viru Mining Region	Barren heated lands (bare mining surfaces)	Waterlogged surfaces, possibly flooded pits

3.3.1 Standard Bivariate Raster Maps

The standard bivariate raster maps for the Tartu study area (**Figure 3-3**) visualize the spatial distribution of PC1 and PC2 values using two classification methods: quantile and equal interval. Similar maps for the other study areas (Tallinn, Porijõgi Catchment, and Ida-Viru Mining Region) are provided in the appendix (**Figure A-4** to **Figure A-6**).

The quantile-based maps display a balanced distribution of colors, effectively highlighting spatial heterogeneity. This method ensures that each color class contains an equal number of data points, leading to a more even representation of PC1 and PC2 combinations across the landscape, regardless of the underlying data distribution. As a result, areas with extreme values (e.g., high PC1/low PC2 or low PC1/high PC2) are clearly visible. In Tartu, quantile-based maps show obvious ecological contrasts. Urban areas with thermal stress are

highlighted in red (high PC2, low PC1), and vegetation zones are highlighted in cyan (low PC2, high PC1).

The equal interval maps, on the other hand, display a visually clustered pattern, with a dominant representation of high PC2 values (red shades) across the map. Because the equal interval method divides the entire range of PC1 and PC2 values into fixed intervals without considering the actual distribution of data. Thus, data values that are densely concentrated in the upper PC2 range (high temperatures) are grouped into the same color class, while lower PC2 values (e.g., wet, vegetated areas) are sparsely represented, leading to their under-visualization.

On the Tartu map, this effect is evident as the warm (red) colors dominate, while the blue-cyan shades representing moist, vegetated areas (low PC2) are almost absent. This visual dominance of high PC2 areas in the equal interval map does not accurately reflect the diversity of ecological conditions but instead amplifies most of the data concentrated in the higher PC2 range. The method's focus on equal numerical intervals leads to an underrepresentation of certain environmental conditions, particularly those with low PC2 values.

The comparison between quantile and equal interval classification reveals a fundamental trade-off:

1. Quantile maps excel at highlighting environmental contrasts and spatial heterogeneity, making them suitable for exploratory analysis and identifying ecological outliers.
2. Equal interval maps, in contrast, provide a direct representation of the PC value ranges, but when data is unevenly distributed, this can lead to a visual dominance of frequently occurring values, potentially obscuring subtle variations within those concentrated ranges and underrepresenting less frequent values, making them less effective for detecting subtle environmental gradients or less common ecological conditions.

The standard bivariate choropleth maps provide a balanced overview of the full variable space but often fail to guide the reader's attention to specific patterns, especially when color categories cluster near the center of the legend.

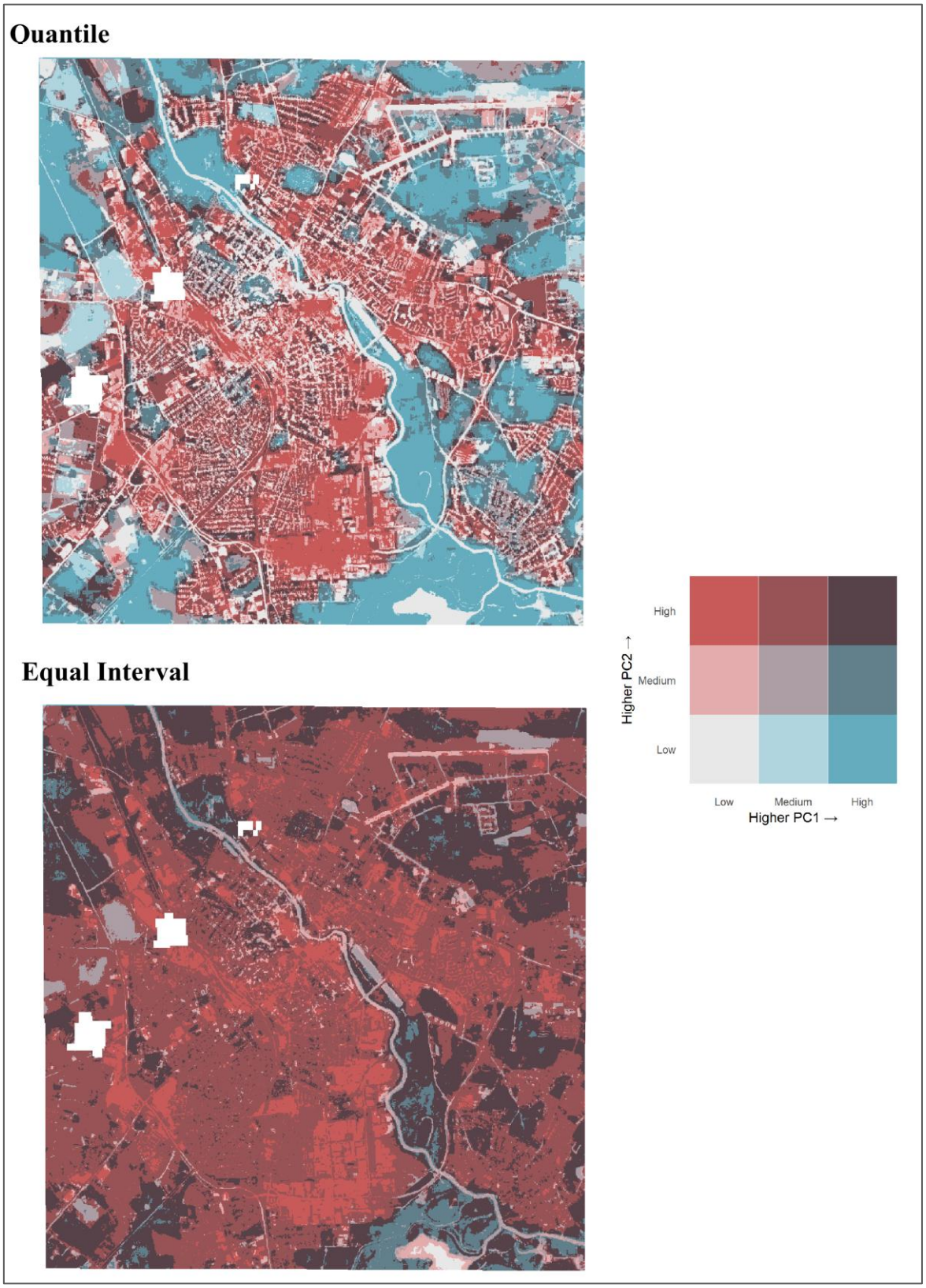


Figure 3-3 Standard bivariate raster maps of Tartu.

3.3.2 Corner Model Maps for Extreme Emphasis

The corner model visualization results for the Tartu study area (**Figure 3-4**) highlight areas exhibiting extreme environmental combinations, prioritizing visual differentiation at the four corners of the legend. Using complementary colors strategically placed at the legend's extreme, this approach enhances perceptual salience of specific environmental contrasts, particularly between dense vegetation with high moisture (green), urban areas with significant thermal stress (magenta), and high-temperature vegetated zones (orange).

Consistent with previous observations (**Section 3.3.1**), the quantile-based classification effectively enhances these visual contrasts, clearly delineating spatial heterogeneity. Extreme value combinations appear more frequently and are better distributed. Conversely, equal interval classification reduces perceptual differentiation, visually compressing the distinctions between ecological extremes.

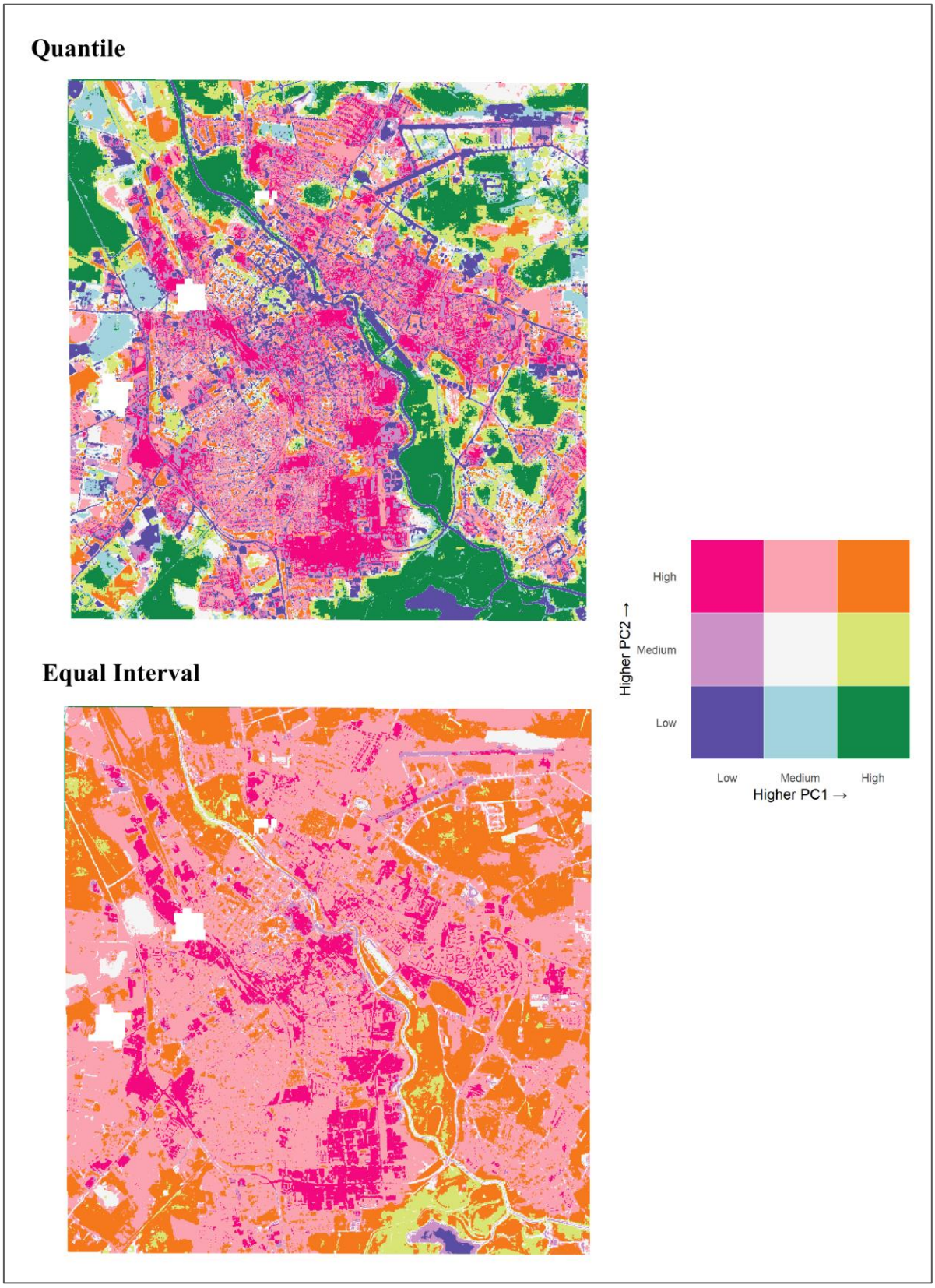


Figure 3-4 Corners model applied to Tartu.

3.3.3 Diagonal Model Maps

Figure 3-5 uses the diagonal model to highlight areas where PC1 and PC2 rise or fall together (i.e., positive correlation). This model draws attention to spatial consistency in environmental gradients, such as urban expansion with thermal intensification or vegetated areas with high moisture retention. The lightest (white) and darkest (black) diagonal classes represent low-low and high-high combinations where correspond to areas previously interpreted in **Table 3-1** as “water bodies, bareness surface but low temperature” and “vegetated urban zones with warm microclimates”.

The gray sequence along the main diagonal indicates the areas where the values of PC1 and PC2 are relatively consistent, suggesting the synergistic changes of these two environmental factors. The dark grey areas are mainly distributed on the urban fringes and some green spaces, representing regions with high vegetation coverage and surface temperatures. The white areas are highly consistent with rivers and small water bodies, indicating areas with high NDWI and low surface temperatures. The gray level in the middle reflects the landscape where vegetation coverage, moisture and surface temperature are all at a medium level.

The blue and orange tones deviating from the main diagonal highlight the areas with significant differences in the PC1 and PC2, indicating the advantage of one environmental factor over the other. The blue tones are mainly concentrated in the urban central area, representing regions with relatively high surface temperatures and low vegetation coverage and moisture. On the contrary, the orange tone is more prominent in the urban periphery and vegetated areas. The deep orange areas indicate regions with high vegetation coverage and high moisture but low surface temperatures, which may correspond to dense forests or shady vegetation.

The diagonal model provides a unique perspective for exploring the spatial relationship between vegetation/moisture and surface temperature in the Tartu region. It clearly shows the synergistic changes and dominant areas of these environmental factors through color coding.

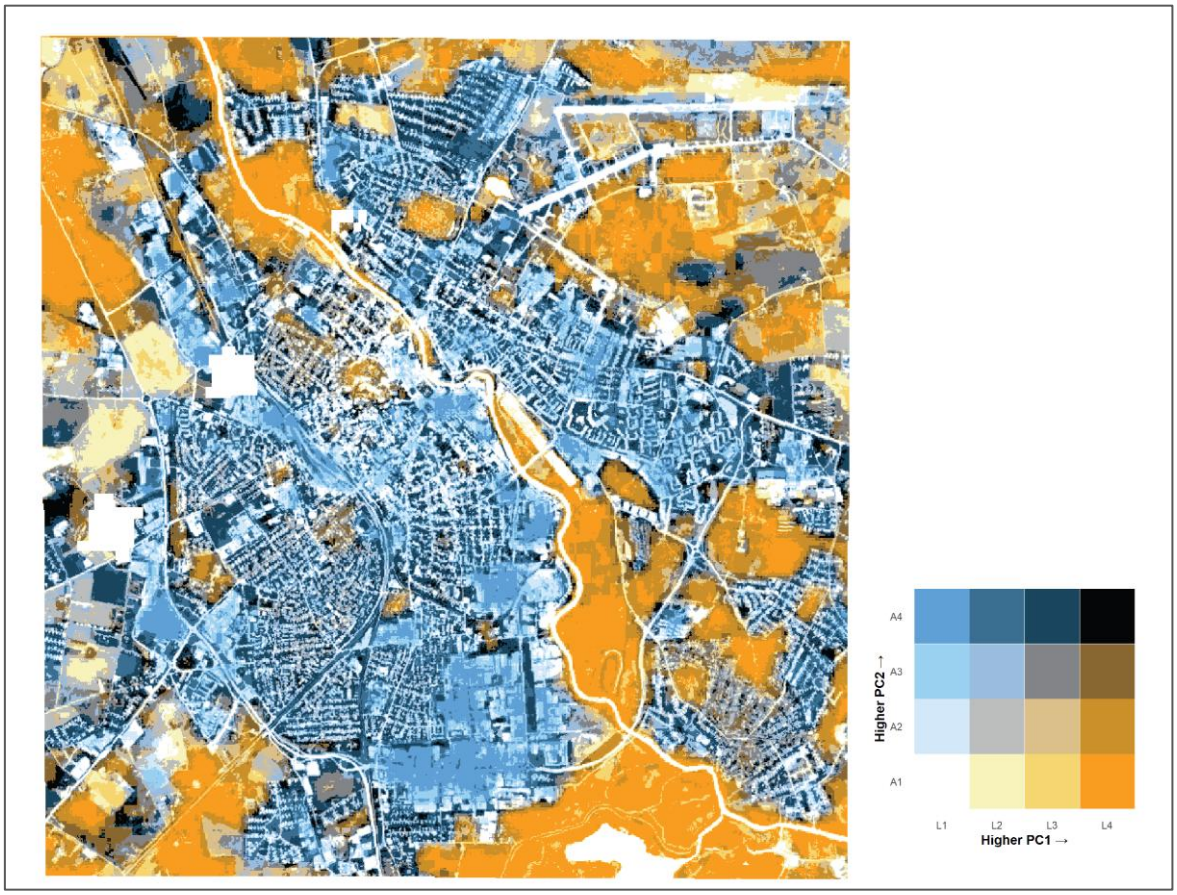


Figure 3-5 Diagonal Model applied to Tartu.

3.3.4 Value-by-alpha Maps for Visual Prioritization

The value-by-alpha model visualizes PCA results in Tartu by encoding PC1 with color (horizontal axis) and PC2 with transparency (vertical axis).**Figure 3-6** shows value-by-alpha map without background, **Figure 3-7** used black background as the base color, while **Figure 3-8** utilizes a white background. The maps for the other study areas (Tallinn, Porijõgi Catchment, and Ida-Viru Mining Region) are provided in the appendix (**Figure A-13** to **Figure A-21**).

The black background (**Figure 3-7**) emphasizes extreme values through high visual contrast. Urban areas with high thermal stress (high PC2, moderate-to-low PC1) are prominently highlighted in saturated reddish orange. Conversely, moist, vegetated zones (higher PC1, lower PC2) appear in darker, opaque blue-purple tones, creating a "spotlighting" effect for high PC2 regions. However, this approach unintentionally diminishes the visibility of regions characterized by high PC1 but low PC2 (e.g., category A1-L5), as these fade into the dark background due to high transparency. Moderate or transitional areas similarly recede visually, further emphasizing ecological extremes at the cost of intermediate gradient visibility.

The white background (**Figure 3-8**) enhances spatial continuity and subtle ecological gradients. Urban areas remain clearly distinguishable in deep orange tones, while vegetated zones appear lighter and more diffuse in blue-purple hues. This version effectively communicates intermediate ecological variations but with reduced visual contrast compared to the black-background alternative.

In summary, while the black-background variant effectively "spotlights" areas with high PC2 values through increased contrast, it sacrifices the visibility of regions representing high PC1 and low PC2 combinations. The white background, by contrast, provides balanced visibility of intermediate environmental conditions, improving overall spatial readability at the expense of visual drama.

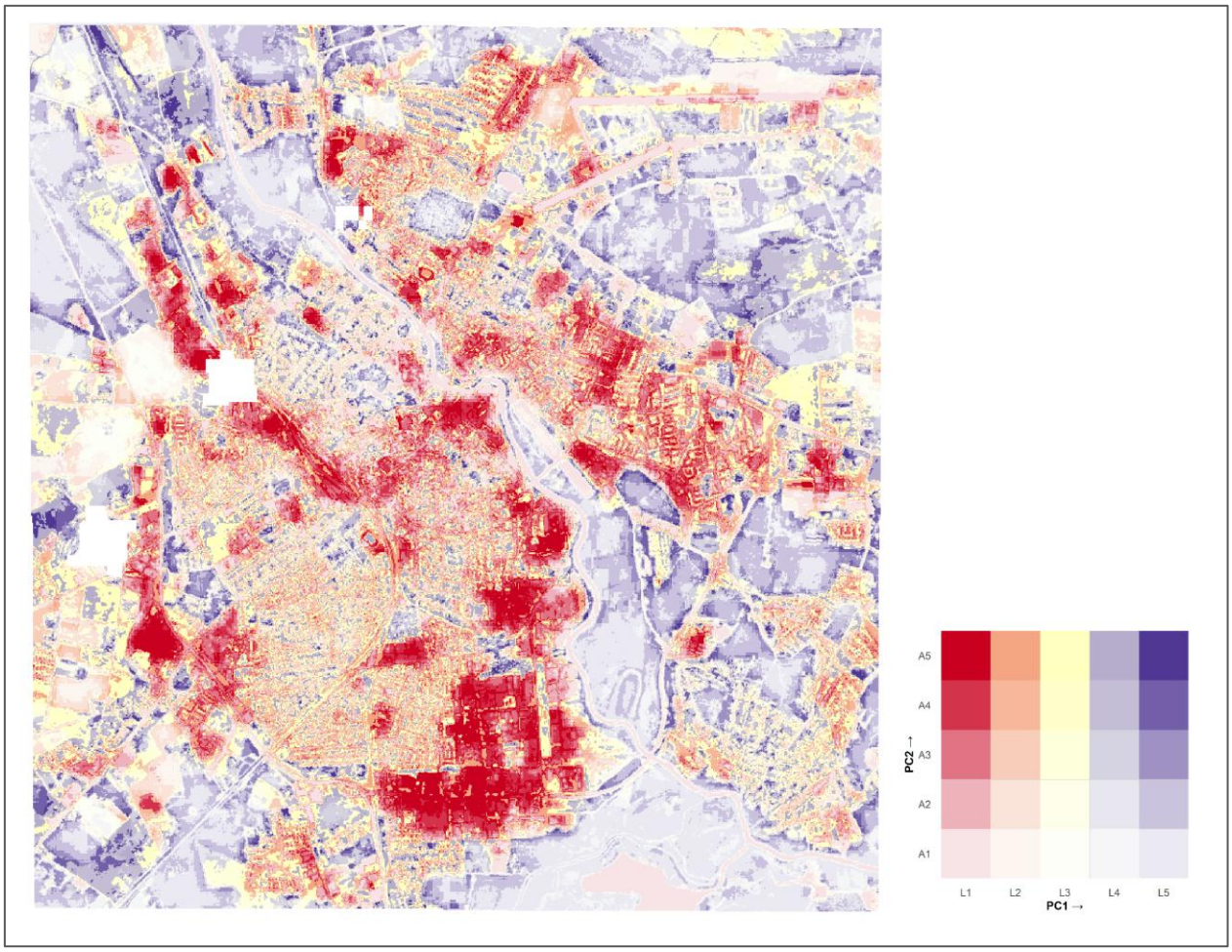


Figure 3-6 Value-by-alpha map of Tartu without background.

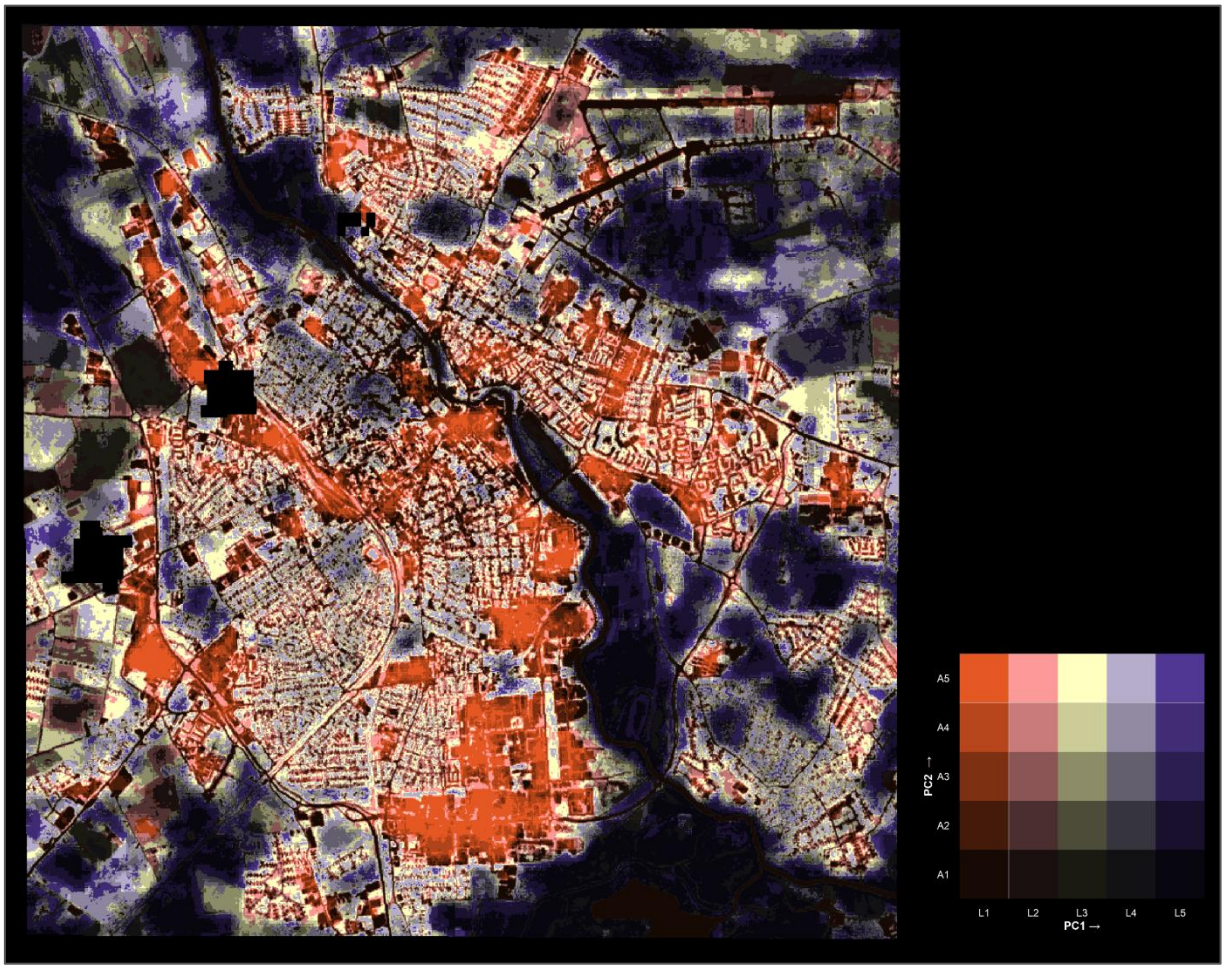


Figure 3-7 Black background value-by-alpha spotlight maps of Tartu.

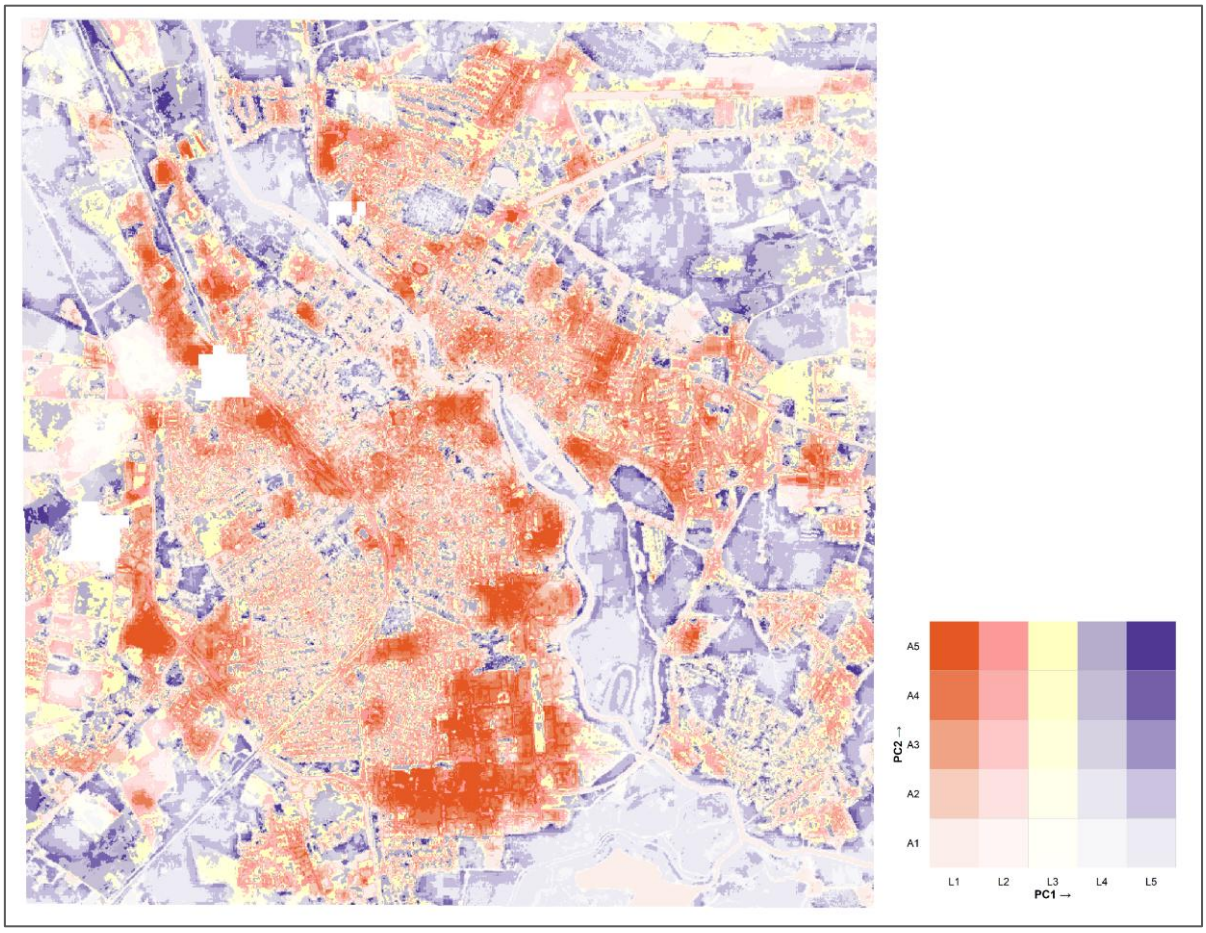


Figure 3-8 White background value-by-alpha spotlight maps of Tartu.

4 Discussion

Univariate and bivariate mapping methods have their own advantages and limitations in data presentation. Univariate maps offer clear and interpretable visualizations of individual environmental indicators. Their categorical classification (e.g., Very High to Low) facilitates direct interpretation, for instance, identifying zones of heat vulnerability or low vegetation with precision and clarity. Despite their clarity, univariate maps are inherently limited by their focus on individual variables, making it challenging to capture interactions between indices or assess composite environmental conditions. The GIS overlay method partially mitigates this by integrating multiple indices into a single composite map, but this approach remains fundamentally constrained. The combined map represents the weighted sum of individual indices, based primarily on PC1 loadings, which may emphasize dominant variables while downplaying others. In contrast, bivariate and alpha-based mapping methods overcome these limitations by encoding two dimensions (PC1 and PC2) in a single visualization. These maps effectively capture the dual aspects of ecological variability, such as vegetation quality (PC1) and high temperature stress (PC2) which reveal a more nuanced environmental pattern.

4.1 PCA-based Dimensionality Reduction for Enhanced Visualization

A key finding of this study is the environmental gradients the environmental gradients captured by PC1 align with previous research, where NDVI (greenness) and NDMI (moisture) consistently show positive loadings, while LST (temperature) and BSI (bare soil) exhibit negative loadings. However, a notable difference emerges in the proportion of variance explained by PC1 compare with similar studies (Liu et al., 2022; Sathyaseelan et al., 2023), PC1 accounts for 75–90% of the total variance, whereas in this study, it explains only 50–80%. This discrepancy can be attributed to two main factors:

1. Index selection: while previous studies often use a limited set of indices (NDVI, NDBSI, LST, LSM), this study incorporates NDVI, BSI, NDWI, NDMI, and LST. The inclusion of NDWI and NDMI introduces additional dimensions of moisture and vegetation conditions.
2. The specific environmental characteristics and spatial heterogeneity of the study area may lead to the first principal component being unable to centrally explain most of the variance as in other studies.

Despite the lower proportion of variance explained by PC1 compared to some previous studies, PCA-based dimensionality reduction significantly enhanced the effectiveness of

multivariate environmental raster visualizations in this research. By transforming five remote sensing indices into uncorrelated principal components, PCA reduced data redundancy and highlighted the underlying environmental gradients.

Specifically, PC1 captured the key vegetation-moisture-dryness gradient, while PC2, being orthogonal to PC1, captured a distinct and crucial temperature gradient that might not be immediately apparent from simple correlation analysis of the raw indices (**Figure 3-1**). For instance, as shown in **Figure 3-4**, specific areas within Tartu show localized patterns of high vegetation co-occurring with elevated surface temperatures (high PC1 and high PC2). These findings demonstrate that PCA not only simplifies complex datasets but also enhances the detection of latent environmental structures, to support more effective and nuanced raster-based visual communication.

4.2 Evaluating Visualization Techniques

Univariate and bivariate mapping methods have their own advantages and limitations in data presentation.

4.2.1 Univariate Mapping

Univariate maps offer clear and interpretable visualizations of individual environmental indicators. Their categorical classification (e.g., Very High to Low) facilitates direct interpretation, for instance, identifying zones of heat vulnerability or low vegetation with precision and clarity. Despite their clarity, univariate maps are inherently limited by their focus on individual variables, making it challenging to capture interactions between indices or assess composite environmental conditions. The GIS overlay method partially mitigates this by integrating multiple indices into a single composite map, but this approach remains fundamentally constrained. The combined map represents the weighted sum of individual indices, based primarily on PC1 loadings, which may emphasize dominant variables while downplaying others.

4.2.2 Bivariate Mapping

In contrast, bivariate and alpha-based mapping methods overcome these limitations by encoding two dimensions (PC1 and PC2) in a single visualization. These maps effectively capture the dual aspects of ecological variability, such as vegetation quality (PC1) and high temperature stress (PC2) which reveal a more nuanced environmental pattern.

- **The corners model:** This model, with its color emphasis on the four extreme corners of the bivariate matrix (**Figure A-8**), significantly improves the salience of critical ecological combinations, such as dense vegetation with high moisture or high temperature stress areas. As Strode et al. (2019) indicate, it is particularly useful for exploring the intersection of extreme values for two variables. Its high-contrast color scheme offers improved visual clarity and differentiation, even in intermediate areas, which can benefit users with color vision deficiencies. However, this strength can also lead to visual clutter in highly diverse landscapes like the Porijõgi Catchment (Figure A-8), potentially causing confusion if users do not clearly identify the desired PC1 and PC2 value ranges via the legend. Furthermore, the grayish-white area in the center, representing medium levels for both PC1 and PC2, may signify ecologically sensitive transitional zones that warrant specific attention in environmental management and monitoring, though this interpretation requires further validation.
- **Diagonal Model:** Designed to reveal spatial relationships among variables without presupposing statistical correlations (Strode et al., 2019), the diagonal model effectively differentiates between contrasting areas, such as urban regions (orange hues for high PC2 and low PC1) and surrounding green spaces (blue hues for high PC1 and low PC2). While its 4x4 color matrix provides finer classification than the Corner Model, the continuous color gradient can introduce perceptual challenges. Subtle color differences along the diagonal (e.g., light blue to dark blue) may be difficult to distinguish in complex landscapes or where PC1 and PC2 values are densely clustered, hindering clarity. Moreover, its emphasis on complementary colors, while highlighting urban-vegetation contrast, may oversimplify more detailed ecological transitions, making it difficult to identify areas with medium or mixed features.
- **Value-by-alpha Mapping**: in their spotlight forms, offer a distinct approach to managing visual hierarchy. By encoding one variable (PC1) with color and another (PC2) with alpha (transparency), they effectively filter or “fade” less relevant combinations from visual attention. This method successfully directs focus toward ecologically meaningful gradients, such as urban heat islands (high PC2), particularly when PC2 controls the transparency channel. However, this design introduces trade-offs: extreme values of the color variable can become nearly invisible if paired with low alpha. For instance, highly saturated colors representing critical combinations like healthy vegetated areas with low temperature (high PC1, low PC2) may be suppressed under low

PC2 transparency. This visual omission is most noticeable in dark-background spotlight maps, where saturated colors disappear entirely into the background, potentially obscuring important minority cases. Interestingly, this attenuation of off-axis extremes in alpha maps mirrors the emphasis on the diagonal axis in the diagonal model, albeit through a different mechanism. If the Diagonal Model is reinterpreted metaphorically as a type of alpha emphasis, where proximity to the diagonal implies opacity and distance implies transparency, it can be seen as an implicit value-by-alpha structure. This interpretation aligns with findings by Hilden (2022). Both methods thus engage with the same core challenge in multivariate visualization: how to represent complex combinations without overwhelming the viewer or obscuring critical minority cases.

4.2.3 Enhancing Clarity and Interpretability

To enhance the clarity and interpretability of these multivariate visualizations, our findings highlight several paramount factors:

- **Strategic Data Classification:** The effectiveness of classification schemes depends heavily on the underlying data distribution, especially in raster-based remote sensing indices, which typically exhibit skewed or multimodal distributions rather than normality. In this study, PC1 was left-skewed (**Figure 4-1**), making it more responsive to quantile classification, which distributes values evenly and enhances spatial contrast. By contrast, equal interval classification assumes a uniform distribution and tends to cluster values into a few dominant classes. This effect was evident in the visualization of PC2, where the map became dominated by high PC2 values, rendering combinations such as high PC1 and low PC2 nearly invisible (**Figure 3-4**). This aligns with previous findings (Kebonye et al., 2023; S. Li & Shan, 2022) that equal interval classification often compresses variation, masking important ecological detail, while the quantile classification often improves contrast and reveals spatial heterogeneity.
- **Intuitive Color Scheme Design:** The success of bivariate and alpha-based maps heavily relies on perceptually uniform and easily distinguishable color palettes. Our findings highlight challenges with overly subtle color differences in the diagonal model and potential information loss in value-by-alpha maps when saturated colors merge with the background under low alpha. Future designs should prioritize visual clarity over an excessive number of categories, especially in complex environmental contexts.

- **Alignment with Map Objectives:** Crucially, the most effective technique is ultimately determined by the map's purpose. For broad public communication, the simplicity and directness of univariate maps might be preferred. For expert analysis aiming at pattern detection or outlier identification, specialized bivariate methods like the corners model or the focused thematic emphasis of value-by-alpha maps may be more appropriate. The diagonal model remains valuable for exploring variable relationships, especially divergence, which can be useful for detecting environmental outliers or mismatched ecological states.

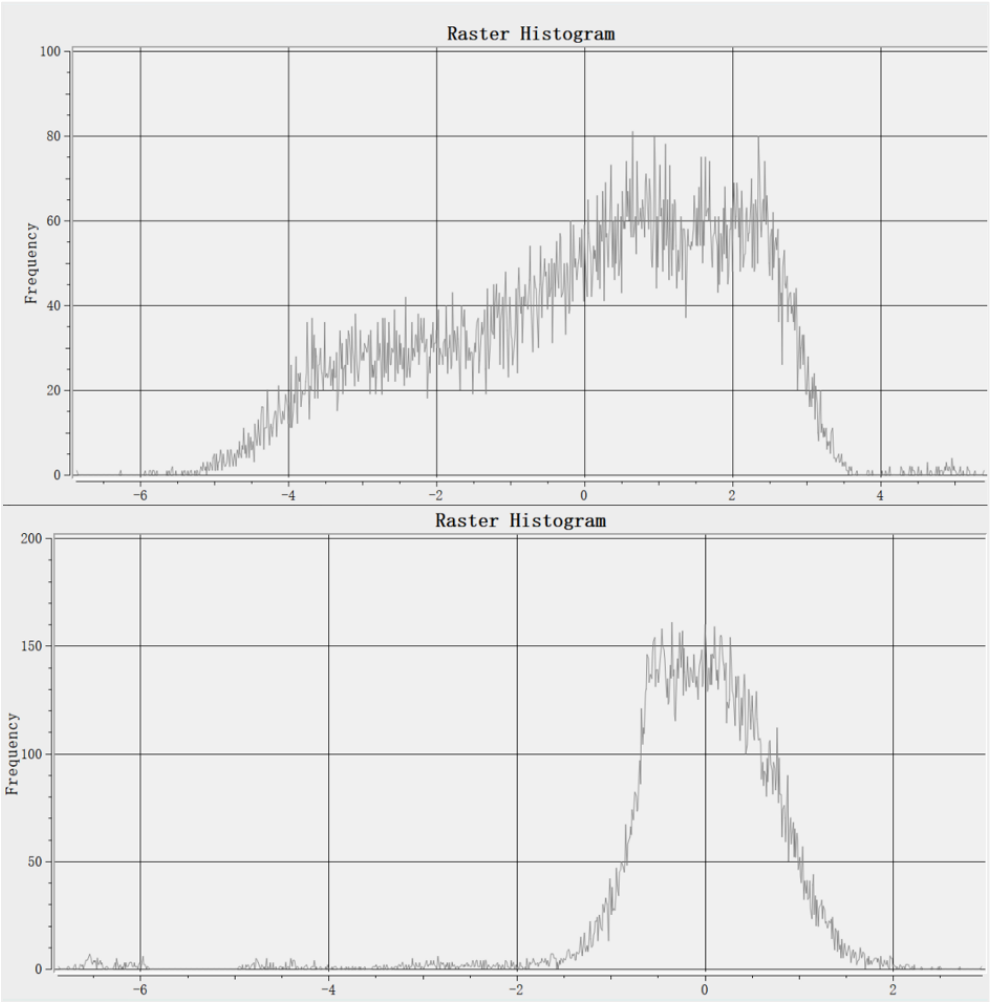


Figure 4-1 Distribution of PC1(Top) and PC2 (bottom), Tartu

This comparative framework underscores that optimizing raster-based visualization techniques requires thoughtful consideration of PCA's role in dimensionality reduction, the inherent strengths and limitations of various cartographic methods, and a strategic alignment with the map's intended message and audience. The selection of visualization method should be guided by its purpose whether the goal is pattern detection, outlier identification, or

focused thematic emphasis to support more effective ecological mapping and informed decision-making.

5 Conclusion

This study systematically assessed the effectiveness of various cartographic visualization techniques: univariate, bivariate, and value-by-alpha mapping, in representing complex multivariate environmental raster data derived from remote sensing indices across diverse Estonian landscapes. Our findings provide a comprehensive framework for optimizing such visualizations, directly addressing how PCA enhances effectiveness, and which techniques are most effective for communication.

To answer the first research question: PCA-based dimensionality reduction proved crucial for enhancing visualization effectiveness by transforming muti-dimensional, correlated remote sensing indices into a more manageable set of uncorrelated principal components. This process not only reduced data redundancy but also distilled complex environmental information into distinct, ecologically meaningful gradients (e.g., PC1 representing vegetation-moisture and PC2 representing temperature), which are not immediately evident from raw index correlations. This demonstrates PCA's ability to reveal latent environmental structures, thereby providing a more nuanced foundation for visual communication.

To answer the second research question: The effectiveness of specific visualization techniques varies significantly depending on the communication objective. Univariate maps, while offering unparalleled clarity for individual indicators and being highly suitable for broad public communication, are inherently limited in capturing multi-dimensional ecological interactions. In contrast, bivariate maps, leveraging the first two principal components, provide a richer depiction of ecological variability. The corners model is particularly effective for highlighting extreme environmental combinations, aiding in outlier detection. The diagonal model, despite perceptual challenges with subtle color gradients, effectively emphasizes divergence between variables, useful for exploring contrasting ecological states. Furthermore, value-by-alpha maps offer a unique ability to guide viewer attention to specific, ecologically meaningful patterns by dynamically adjusting transparency, proving valuable for thematic emphasis (e.g., high temperature stress areas). To enhance the clarity and interpretability of these visualizations, several factors are paramount. Strategic data classification, such as using quantile methods for skewed data, significantly improves visual contrast and reveals spatial heterogeneity. Intuitive color scheme design, prioritizing perceptual uniform and distinguishable palettes, is essential for accurate interpretation. Crucially, the alignment of the chosen visualization method with the map's objective whether

for exploratory analysis, public outreach, or expert interpretation is fundamental for effective communication.

These findings have several implications. Firstly, this research advances our understanding of PCA's role in enhancing multivariate visualization effectiveness. We demonstrate how PCA, beyond simple dimensionality reduction, critically reveals underlying, orthogonal ecological gradients and their interactions (e.g., PC1 capturing vegetation-moisture vs. PC2 for temperature, as seen in Figure 3-1), providing a richer and more interpretable foundation for visual communication. Secondly, this study provides direct and practical guidance for environmental cartography practices. Our systematic comparison offers empirical insights for selecting and optimizing visualization techniques and classification methods based on specific ecological questions and communication objectives. This empowers professionals to create more strategic and effective maps for informed decision-making in ecological assessment and conservation.

This study, while providing a comprehensive framework for optimizing environmental raster visualizations, still has some limitations. Firstly, temporal and spatial constraints limit the generalizability of our findings, as the analysis relied solely on summer 2024 data and was confined to localized areas in Estonia. Future research should incorporate multi-seasonal and broader-scale datasets to validate these techniques across varying environmental conditions and geographical extents. Secondly, the absence of dedicated user evaluation or feedback meant our assessment of visual effectiveness was based on objective analysis, lacking insights into subjective perceptual strengths and weaknesses. Subsequent studies should integrate user-centric evaluations to refine design principles for enhanced communication.

6 Summary

Comparing raster visualization techniques for environmental indices: univariate and bivariate mapping in Estonia

Wenyi Fang

Summary

Environmental data visualization is essential for understanding ecological conditions and informing sustainable management. Traditional single-variable mapping methods, such as GIS-based weighted overlay maps, provide a clear but limited view of environmental patterns by isolating individual indices (e.g., NDVI, NDWI, LST, and BSI). However, such univariate approaches cannot capture the complex interactions between multiple ecological factors. This study addresses this limitation by exploring raster visualization techniques that integrate multiple environmental indices using Principal Component Analysis (PCA) and various bivariate mapping methods. Specifically, the following two research questions were addressed:

1. How can PCA-based dimensionality reduction enhance the effectiveness of multivariate environmental raster visualizations?
2. What are the most effective visualization techniques for communicating multivariate environmental raster data, and how can their clarity and interpretability be enhanced?

To answer these questions, this study applied three distinct visualization techniques across four diverse regions in Estonia (Tallinn, Tartu, Porijõgi, and the Ida-Viru Mining Region):

1. GIS-based weighted overlay maps: Using the first principal components (PC1) from PCA as the weighting basis, as a univariate map comparison.
2. Bivariate color maps: Visualize the first two principal components (PC1 and PC2) using classification schemes (quantile, equal interval) and visual models (corners, diagonal).
3. Value-by-alpha maps: Combine color and transparency to encode PC1 and PC2, respectively.

The findings revealed that PCA-based dimensionality reduction provided a crucial foundation for effective multivariate environmental raster visualizations by condensing complex, high-dimensional datasets into a more manageable number of uncorrelated principal components, primarily PC1 and PC2. This simplification allowed for the creation

of intuitive bivariate and alpha-based maps that could reveal nuanced ecological interactions often obscured in raw multi-band imagery.

The GIS overlay offered clear and intuitive representations of ecological sustainability, particularly suitable for non-expert audiences. However, their reliance solely on PC1, which explained a significant but limited portion of the total variance (60-80%), tended to oversimplify underlying ecological complexity and potentially obscure important secondary gradients captured by other PCs. By contrast, the bivariate color maps, leveraging the information from both PC1 and PC2, captured more nuanced ecological interactions, with quantile classification providing greater contrast and visual differentiation crucial for detailed analysis compared to the more uniform representation of equal interval classification. Among the bivariate methods, the corners model effectively highlighted critical ecological combinations representing extreme environmental conditions, while the diagonal model, though intended to emphasize variable relationships, primarily differentiated areas based on high/low combinations of PC1 and PC2. Finally, the value-by-alpha maps demonstrated a unique ability to focus viewer attention on critical ecological patterns such as urban heat islands (high PC2) and vegetated areas with high moisture content (high PC1, low PC2) by adjusting transparency based on one variable. However, this method required careful consideration to avoid obscuring important information associated with extreme values of the color-encoded variable under low alpha.

In conclusion, the effectiveness of multivariate environmental raster visualization is shaped by the ability to balance information richness with clarity and interpretability. While simpler methods like weighted overlays offer ease understanding, they may oversimplify complex ecological realities. Bivariate and alpha-based methods can reveal more nuanced patterns, but their effectiveness relies heavily on thoughtful color scheme design, appropriate data classification, and clear legend creation. For communicating complex ecological data derived from PCA, techniques that allow for the visualization of at least the first two principal components, such as well-designed bivariate color maps and value-by-alpha maps, are more effective in capturing the multifaceted nature of environmental variability. The choice of technique should be guided by the specific research question and the need to highlight aspects of the data, whether extreme conditions, variable relationships, or the dominance of specific ecological gradients.

Keskkonnaindeksite rastervisualiseerimise tehnikate võrdlus: ühe- ja kahemõõtmeline kaardistamine Eestis

Wenyi Fang

Kokkuvõte

Keskkonnaandmete visualiseerimine on oluline ökoloogiliste tingimuste mõistmiseks ja säästva majandamise toetamiseks. Traditsioonilised ühemõõtmelised kaardistamismeetodid, nagu GIS-põhised kaalutud ülekattekaardid, pakuvad selget, kuid piiratud vaadet keskkonnamustritele, keskendudes üksikutele indeksitele (nt NDVI, NDWI, LST ja BSI). Sellised ühemõõtmelised lähenemised ei suuda aga haarata mitme ökoloogilise teguri keerukaid vastastikmõjusid. Käesolev töö uurib rastervisualiseerimise tehnikaid, mis integreerivad mitmeid keskkonnaindekseid, kasutades peakomponentide analüüsi (PCA) ja erinevaid kahemõõtmelisi kaardistamismeetodeid. Täpsemalt käsitleti kahte uurimisküsimust:

1. Kuidas saab PCA-põhine dimensioonide vähendamine parandada mitmemõõtmeliste keskkonnaandmete rastervisualiseerimise tõhusust?
2. Millised on kõige tõhusamad visualiseerimistehnikad mitmemõõtmeliste keskkonnaandmete edastamiseks ja kuidas saab parandada nende selgust ja tõlgendatavust?

Nendele küsimustele vastamiseks kasutati kolme erinevat visualiseerimistehnikat neljas erinevas piirkonnas Eestis (Tallinn, Tartu, Porijõgi ja Kaevandusala):

1. GIS-põhised kaalutud ülekattekaardid: Kaardid, mis kasutavad PCA põhjal leitud PC1 kaalutegurina. Esitavad ökoloogilist jätkusuutlikkust ühemõõtmelisena.
2. Kahemõõtmelised värvikaardid: Kaardid, mis kombineerivad PC1 ja PC2, kasutades erinevaid klassifitseerimisskeeme (kvantiil ja võrdsed intervallid) ning värvimudeleid (nurgad ja diagonaalid), et visualiseerida ökoloogilisi vastasmõjusid.
3. Läbipaistvuse põhised kaardid (value-by-alpha): Kaardid, mis kombineerivad PC1 ja PC2, kus üks muutuja on kodeeritud värvina ja teine läbipaistvusega, rõhutades olulisi ökoloogilisi mustreid ja vähendades visuaalset keerukust.

Tulemused näitasid, et PCA-põhine dimensioonide vähendamine pakkus tõhusa keskkonna mitmemõõtmelise rastervisualiseerimise jaoks olulise aluse, koondades keerulised ja kõrgedimensioonilised andmekogumid hallatavamasse arvu omavahel korreleerimata

põhikomponentidesse, peamiselt PC1 ja PC2. See lihtsustamine võimaldas luua intuitiivseid kahemõõtmelisi ja läbipaistvuse põhiseid kaarte, mis suutsid esile tuua ökoloogilisi vastastikmõjusid, mis algsetes multispektraalfotodes sageli varju jäävad.

GIS-põhised kaalutud ülekattekaardid pakkusid selget ja arusaadavat ökoloogilise jätkusuutlikkuse esitust, mis sobisid eriti hästi mitteekspertidest publikule. Siiski kaldusid need liigselt toetuma ainult PC1-le, mis selgitas küll märkimisväärset, kuid piiratud osa koguväärtusest (60–80%), lihtsustades sageli looduse keerukust ja jättes välja teiste peakomponentide poolt leitud olulisi teiseseid gradiente. Kahemõõtmeliselt värvikaartidelt, mis kasutasid nii PC1 kui ka PC2 teavet, võis näha keerukamaid ökoloogilisi vastasmõjusid. Detailse analüüsi jaoks pakkus kvantiiliklassifikatsioon suuremat kontrasti ja visuaalset eristuvust võrreldes võrdsete intervallide meetodiga, mis oli ühtlasem.

Kahemõõtmelistest meetoditest tõi nurkamudel tõhusamalt esile olulised ökoloogilised kombinatsioonid, mis esindasid äärmuslikke keskkonnatingimusi. Diagonaalmudel, mis oli mõeldud muutujate seoste esile toomiseks, eristas aga peamiselt piirkondi, kus PC1 ja PC2 väärtused olid kõrged või madalad.

Läbipaistvuse põhised kaardid (value-by-alpha) näitasid ainulaadset võimet juhtida vaataja tähelepanu olulistele ökoloogilistele mustritele nagu linnasoojusaarte piirkonnad (kõrge PC2) ja kõrge niiskusesisaldusega taimestikualad (kõrge PC1, madal PC2), reguleerides läbipaistvust ühe muutuja alusel. Siiski nõudis see meetod hoolikat ette valmistamist, et mitte peita olulist teavet, mis on seotud värviga kodeeritud muutuja äärmuslike väärtustega madala läbipaistvuse korral.

Kokkuvõttes sõltub mitmemõõtmelise keskkonnateabe rastervisualiseerimise efektiivsus meetodi valikust, et leida tasakaal edasiantava teabe mahu ning selguse ja tõlgendatavuse vahel. Kuigi lihtsamad meetodid, nagu kaalutud ülekattekaardid, on kergemini mõistetavad, võivad need liigselt tegelikku olukorda lihtsustada. Kahemõõtmelised ja läbipaistvuse põhised meetodid toovad detailid paremini esile, kuid nendest arusaadamise jaoks on vaja hästi läbimõeldud värviskeemi, sobivat andmete klassifitseerimist ja selget legendi.

PCA abil saadud keskonnaandmete visualiseerimisel on tõhusaimad meetodid, mis kasutavad vähemalt kahte esimest põhikomponenti, näiteks hästi kujundatud kahemõõtmelised värvikaardid ja läbipaistvuse põhised kaardid. Meetodi valik peaks vastama konkreetsele uurimisküsimusele ja rõhutama aspekte vastavalt vajadusele, olgu selleks äärmuslikud tingimused, muutujate suhted või teatud ökoloogiliste gradientide domineerimine.

Acknowledgment

I would like to thank my supervisor, Dr. Alexander Kmoch, for his continuous support throughout the entire research process. His insightful feedback, constructive suggestions, and engaging discussions greatly contributed to shaping the structure and depth of this thesis. I am especially grateful for his patience and encouragement, which guided me step by step through the complexities of the study, helping me refine my ideas and develop a clearer direction.

I would also like to express my sincere gratitude to Dr. Merle Muru and PhD student Raúl García Estévez from the Department of Geography at the University of Tartu for their valuable reviews and constructive feedback, which significantly improved the quality of this thesis. I am especially thankful to Prof. Evelyn Uuemaa for her guidance on data-related aspects, and to Mr. Joonatan Kama from the Estonian Environment Agency for his insightful input. My thanks also extend to PhD student Xiao Cai (University of Helsinki), Dr. Junming Ke (Zhejiang University), and PhD students Zhiming Ma and Yuyan (University of Helsinki), whose thoughtful suggestions and detailed comments provided both conceptual insight and practical direction for revising this work.

Finally, I would like to thank my family members for their unwavering support and encouragement throughout my studies. Their understanding and care have been a constant source of strength during this academic journey.

References

An, M., Xie, P., He, W., Wang, B., Huang, J., & Khanal, R. (2022). Spatiotemporal change of ecologic environment quality and human interaction factors in three gorges ecologic economic corridor, based on RSEI. *Ecological Indicators*, *141*, 109090. https://doi.org/10.1016/j.ecolind.2022.109090

Annatakarn, K., Annatakarn, K., Fooprateepsiri, R., Suwanprapab, M., Supunyachotsakul, C., & Witchayangkoon, B. (2022). Finding threshold for NDVI to classify green area: Case study in the central Thailand. *Journal of Hunan University Natural Sciences*, *49*(4).

Artikanur, S. D., Widiatmaka, Setiawan, Y., & Marimin. (2022). Normalized Difference Drought Index (NDDI) computation for mapping drought severity in Bojonegoro Regency, East Java, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *1109*(1), 012027. https://doi.org/10.1088/1755-1315/1109/1/012027

Bazaglia Filho, O., Rizzo, R., Lepsch, I. F., Prado, H. do, Gomes, F. H., Mazza, J. A., & Demattê, J. A. M. (2013). Comparison between detailed digital and conventional soil maps of an area with complex geology. *Revista Brasileira de Ciência Do Solo*, *37*.

Boiarskii, B., & Hasegawa, H. (2019). Comparison of NDVI and NDRE indices to detect differences in vegetation and chlorophyll content. *J. Mech. Contin. Math. Sci*, *4*, 20–29.

Cartone, A., & Postiglione, P. (2021). Principal component analysis for geographical data: The role of spatial effects in the definition of composite indicators. *Spatial Economic Analysis*, *16*(2), 126–147.

Chandrasekar, K., & Sesha Sai, M. V. R. (2015). Monitoring of late-season agricultural drought in cotton-growing districts of Andhra Pradesh state, India, using vegetation,

water and soil moisture indices. *Natural Hazards*, *75*(2), 1023–1046. https://doi.org/10.1007/s11069-014-1364-4

Chen, Y., Yu, J., & Khan, S. (2010). Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. *Environmental Modelling & Software*, *25*(12), 1582–1591. https://doi.org/10.1016/j.envsoft.2010.06.001

Cheng, S., Xu, W., & Mueller, K. (2019). ColorMapND: A Data-Driven Approach and Tool for Mapping Multivariate Data to Color. *IEEE Transactions on Visualization and Computer Graphics*, *25*(2), 1361–1377. IEEE Transactions on Visualization and Computer Graphics. https://doi.org/10.1109/TVCG.2018.2808489

Color Scales: Continuous vs Discrete / Chris Henrick | Observable. (n.d.). Retrieved April 9, 2025, from https://observablehq.com/@clhenrick/color-scales-continuous-vs-discrete

Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A. S., & McLoone, S. (2013). Principal component analysis on spatial data: An overview. *Annals of the Association of American Geographers*, *103*(1), 106–128.

Diamond, L. (2019). Vector Formats and Sources. In *The Geographic Information Science & Technology Body of Knowledge (4th Quarter 2019 Edition), John P. Wilson (ed.)*. https://doi.org/10.22224/gistbok/2019.4.8.

Earth Resources Observation and Science (EROS) Center. (2020). *Landsat 8-9 Operational Land Imager / Thermal Infrared Sensor Level-2, Collection 2* [Dataset]. U.S. Geological Survey. https://doi.org/10.5066/P9OGBGM6

Eensaar, A. (2016). Temporal and spatial variability of air temperatures in Estonia during 1756–2014. *Journal of Climatology*, *2016*(1), 9426791.

Elif Bulut & Helen Thompson. (2023, July). Performing principal component analysis (PCA) to determine weights for index indicators. *Esri*. https://www.esri.com/arcgis-

blog/products/api-python/analytics/performing-principal-component-analysis-pca-to-determine-weights-for-index-indicators

EOS DATA ANALYTICS. (2021, September 29). *NDWI: Index Formula, Value Range, And Uses In Agriculture*. https://eos.com/make-an-analysis/ndwi/

Eslamirad, N., De Luca, F., Lylykangas, K. S., Ben Yahia, S., & Rasoulinezhad, M. (2023). Geoprocess of geospatial urban data in Tallinn, Estonia. *Data in Brief*, *48*, 109172. https://doi.org/10.1016/j.dib.2023.109172

Estonian Environment Agency. (2020). *Clinmate normals*. https://www.ilmateenistus.ee/kliima/kliimanormid/sademed/?lang=en

Faisal, K., & Shaker, A. (2017). An Investigation of GIS Overlay and PCA Techniques for Urban Environmental Quality Assessment: A Case Study in Toronto, Ontario, Canada. *Sustainability*, *9*(3), Article 3. https://doi.org/10.3390/su9030380

Ferreira, M. B., Ferreira, R. L. C., da Silva, J. A. A., de Lima, R. B., Silva, E. A., de Sousa, A. N., De La Cruz, D. B. C., & da Silva, M. V. (2024). Spatial-Temporal Dynamics of Water Resources in Seasonally Dry Tropical Forest: Causes and Vegetation Response. *AgriEngineering*, *6*(3), Article 3. https://doi.org/10.3390/agriengineering6030148

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Golebiowska, I., Korycka-Skorupa, J., Slomska-Przech, K., & Wilson, J. P. (2021). Common Thematic Map Types. In *The Geographic Information Science & Technology Body of Knowledge*. https://doi.org/10.22224/gistbok/2021.2.7

Hilden, J. (2022). *Bivariate hue blending–Color scale design for bivariate choropleth maps with a custom tool*.

Hu, X., & Xu, H. (2018). A new remote sensing index for assessing the spatial heterogeneity in urban ecological quality: A case from Fuzhou City, China. *Ecological Indicators*, *89*, 11–21. https://doi.org/10.1016/j.ecolind.2018.02.006

Ibraheem, N. A., Hasan, M. M., Khan, R. Z., & Mishra, P. K. (2012). Understanding color models: A review. *ARPN Journal of Science and Technology*, *2*(3), 265–275.

Ji, L., Zhang, L., & Wylie, B. (2009). Analysis of Dynamic Thresholds for the Normalized Difference Water Index. *Photogrammetric Engineering & Remote Sensing*, *75*(11), 1307–1317. https://doi.org/10.14358/PERS.75.11.1307

Jin, S., & Sader, S. A. (2005). Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sensing of Environment*, *94*(3), 364–372. https://doi.org/10.1016/j.rse.2004.10.012

Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.

Joshua Stevens. (2015, February 18). *Bivariate Choropleth Maps: A How-to Guide*. https://www.joshuastevens.net/cartography/make-a-bivariate-choropleth-map/

Kebonye, N. M., Agyeman, P. C., Seletlo, Z., & Eze, P. N. (2023). On exploring bivariate and trivariate maps as visualization tools for spatial associations in digital soil mapping: A focus on soil properties. *Precision Agriculture*, *24*(2), 511–532. https://doi.org/10.1007/s11119-022-09955-7

Kebonye, N. M., Eze, P. N., John, K., Gholizadeh, A., Dajčl, J., Drábek, O., Němeček, K., & Borůvka, L. (2021). Self-organizing map artificial neural networks and sequential Gaussian simulation technique for mapping potentially toxic element hotspots in polluted mining soils. *Journal of Geochemical Exploration*, *222*, 106680. https://doi.org/10.1016/j.gexplo.2020.106680

Laroche, V. (2019). Making urban stormwater management more sustainable. A case study of Tallinn, Estonia. *IIIEE Master Thesis*.

Leonowicz, A. (2006). Two-variable choropleth maps as a useful tool for visualization of geographical relationship. *Geografija*, *42*, 33–37.

Li, J., Pei, Y., Zhao, S., Xiao, R., Sang, X., & Zhang, C. (2020). A Review of Remote Sensing for Environmental Monitoring in China. *Remote Sensing*, *12*(7), Article 7. https://doi.org/10.3390/rs12071130

Li, S., & Shan, J. (2022). Adaptive geometric interval classifier. *ISPRS International Journal of Geo-Information*, *11*(8), 430.

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I. F., & Sobrino, J. A. (2013). Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, *131*, 14–37. https://doi.org/10.1016/j.rse.2012.12.008

Liao, W., & Jiang, W. (2020). Evaluation of the Spatiotemporal Variations in the Eco-environmental Quality in China Based on the Remote Sensing Ecological Index. *Remote Sensing*, *12*(15). https://doi.org/10.3390/rs12152462

Liu, Q., Yu, F., & Mu, X. (2022). Evaluation of the Ecological Environment Quality of the Kuye River Source Basin Using the Remote Sensing Ecological Index. *International Journal of Environmental Research and Public Health*, *19*(19), Article 19. https://doi.org/10.3390/ijerph191912500

Lucchesi, L. R., & Wikle, C. K. (2017). Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat*, *6*(1), 292–302. https://doi.org/10.1002/sta4.150

Maharjan, B., Pachel, K., & Loigu, E. (2016). TRENDS IN URBAN STORM WATER QUALITY IN TALLINN AND INFLUENCES FROM STORMFLOW AND BASEFLOW. *Journal of Water Security (JWS)*, *2*.

Malakhov, D. V., & Tsychuyeva, N. Yu. (2020). Calculation of the biophysical parameters of vegetation in an arid area of south-eastern Kazakhstan using the normalized

difference moisture index (NDMI). *Central Asian Journal of Environmental Science and Technology Innovation*, *1*(4), 189–198. https://doi.org/10.22034/CAJESTI.2020.04.01

McFEETERS, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

McFeeters, S. K. (2013). Using the Normalized Difference Water Index (NDWI) within a Geographic Information System to Detect Swimming Pools for Mosquito Abatement: A Practical Approach. *Remote Sensing*, *5*(7), 3544–3561. https://doi.org/10.3390/rs5073544

Messer, L. C., Jagai, J. S., Rappazzo, K. M., & Lobdell, D. T. (2014). Construction of an environmental quality index for public health research. *Environmental Health*, *13*(1), 39. https://doi.org/10.1186/1476-069X-13-39

Moges, D. M., Kmoch, A., & Uuemaa, E. (2022). Application of satellite and reanalysis precipitation products for hydrological modeling in the data-scarce Porijõgi catchment, Estonia. *Journal of Hydrology: Regional Studies*, *41*, 101070. https://doi.org/10.1016/j.ejrh.2022.101070

Mzid, N., Pignatti, S., Huang, W., & Casa, R. (2021). An analysis of bare soil occurrence in arable croplands for remote sensing topsoil applications. *Remote Sensing*, *13*(3), 474.

Nelson, J., & Wilson, J. P. (2020). Multivariate Mapping. In *The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2020 Edition)*. https://doi.org/10.22224/gistbok/2020.1.5

Neto, J. H. (2022). bivariatemaps: An R Package to Create Bivariate Maps and Grid-Intersected Shapes Based on Area Coverage. *Authorea Preprints*.

Nguyen, C. T., Chidthaisong, A., Kieu Diem, P., & Huo, L.-Z. (2021). A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. *Land*, *10*(3). https://doi.org/10.3390/land10030231

Nichol, J., & Wong, M. S. (2009). Mapping Urban Environmental Quality Using Satellite Data and Multiple Parameters. *Environment and Planning B: Planning and Design*, *36*(1), 170–185. https://doi.org/10.1068/b34034

Ning, L., Jiayao, W., & Fen, Q. (2020). The improvement of ecological environment index model RSEI. *Arabian Journal of Geosciences*, *13*(11), 403. https://doi.org/10.1007/s12517-020-05414-7

Palanisamy, P. A., Jain, K., & Bonafoni, S. (2023). Machine Learning Classifier Evaluation for Different Input Combinations: A Case Study with Landsat 9 and Sentinel-2 Data. *Remote Sensing*, *15*(13), Article 13. https://doi.org/10.3390/rs15133241

Panahi, H., Azizi, Z., Kiadaliri, H., Almodaresi, S. A., & Aghamohamadi, H. (2024). Bare soil detecting algorithms in western iran woodlands using remote sensing. *Smart Agricultural Technology*, *7*, 100429. https://doi.org/10.1016/j.atech.2024.100429

Patil, P. P., Jagtap, M. P., Khatri, N., Madan, H., Vadduri, A. A., & Patodia, T. (2024). Exploration and advancement of NDDI leveraging NDVI and NDWI in Indian semi-arid regions: A remote sensing-based study. *Case Studies in Chemical and Environmental Engineering*, *9*, 100573. https://doi.org/10.1016/j.cscee.2023.100573

Persson, M. (2020). *A survey of methods for visualizing spatio-temporal data*.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. Chr. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, *20*(9), 503–510. https://doi.org/10.1016/j.tree.2005.05.011

Porter, T., & Duff, T. (1984). Compositing digital images. *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, 253–259. https://doi.org/10.1145/800031.808606

Racoviteanu, A., Kazak, A., Secrieru, N., & Ivanovici, M. (2024). Fuzzy Logic-Based Visualization and Interpretation of NDVI Maps. *2024 14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–5. https://doi.org/10.1109/WHISPERS65427.2024.10876428

Roth, R. E., Woodruff, A. W., & Johnson, Z. F. (2010). Value-by-alpha maps: An alternative technique to the cartogram. *The Cartographic Journal*, *47*(2), 130–140. https://doi.org/10.1179/000870409X12488753453372

Sarkar, D., Saha, S., & Mondal, P. (2023). Modelling agricultural land suitability for vegetable crops farming using RS and GIS in conjunction with bivariate techniques in the Uttar Dinajpur district of Eastern India. *Green Technologies and Sustainability*, *1*(2), 100022. https://doi.org/10.1016/j.grets.2023.100022

Sathyaseelan, M., Ghosh, S. K., & Ojha, C. S. P. (2023). ENVIRONMENTAL SUSTAINABILITY ASSESSMENT OF A HIMALAYAN CATCHMENT WITH LAND COVER INDICES AND LST RELATIONSHIP USING PRINCIPAL COMPONENT ANALYSIS – A GEOSPATIAL APPROACH. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, *XLVIII-M-1–2023*, 285–292. https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-285-2023

Sentinel Hub. (2024). *NDWI (Normalized Difference Water Index)* [Computer software]. GitHub. https://github.com/sentinel-hub/custom-scripts/blob/main/sentinel-2/ndwi/README.md

Shah, S. A. A., Zhou, P., Walasai, G. D., & Mohsin, M. (2019). Energy security and environmental sustainability index of South Asian countries: A composite index

approach. *Ecological Indicators*, *106*, 105507. https://doi.org/10.1016/j.ecolind.2019.105507

Sharma, R., & Joshi, P. K. (2016). Mapping environmental impacts of rapid urbanization in the National Capital Region of India using remote sensing inputs. *Urban Climate*, *15*, 70–82. https://doi.org/10.1016/j.uclim.2016.01.004

Stachoň, Z., Čeněk ,Jiří, Lacko ,David, Havelková ,Lenka, Hanus ,Martin, Lu ,Wei-Lun, Šašinková ,Alžběta, Ugwitz ,Pavel, Shen ,Jie, & and Šašinka, Č. (2025). A comparison of performance using extrinsic and intrinsic bivariate cartographic visualizations with respect to cognitive style in experienced map users. *Cartography and Geographic Information Science*, *52*(1), 55–68. https://doi.org/10.1080/15230406.2023.2264752

Strode, G., Morgan, J. D., Thornton, B., Mesev, V., Rau, E., Shortes, S., & Johnson, N. (2019). Operationalizing Trumbo's Principles of Bivariate Choropleth Map Design. *Cartographic Perspectives*, *94*, Article 94. https://doi.org/10.14714/CP94.1538

Taloor, A. K., Drinder Singh Manhas, & Chandra Kothyari, G. (2021). Retrieval of land surface temperature, normalized difference moisture index, normalized difference water index of the Ravi basin using Landsat data. *Applied Computing and Geosciences*, *9*, 100051. https://doi.org/10.1016/j.acags.2020.100051

Teffera, Z. L., Li, J., Debsu, T. M., & Menegesha, B. Y. (2018). Assessing land use and land cover dynamics using composites of spectral indices and principal component analysis: A case study in middle Awash subbasin, Ethiopia. *Applied Geography*, *96*, 109–129. https://doi.org/10.1016/j.apgeog.2018.05.015

Teuling, A. J., Stöckli, R., & Seneviratne, S. I. (2011). Bivariate colour maps for visualizing climate data. *International Journal of Climatology*, *31*(9), 1408–1412. https://doi.org/10.1002/joc.2153

Trumbo, B. E. (1981). A Theory for Coloring Bivariate Statistical Maps. *The American Statistician*, *35*(4), 220–226. https://doi.org/10.1080/00031305.1981.10479360

W. -t. Cai, Y. -x. Liu, M. -c. Li, Y. Zhang, & Z. Li. (2010). A best-first multivariate decision tree method used for urban land cover classification. *2010 18th International Conference on Geoinformatics*, 1–5. https://doi.org/10.1109/GEOINFORMATICS.2010.5567871

Wang, Z., Chen, T., Zhu, D., Jia, K., & Plaza, A. (2023). RSEIFE: A new remote sensing ecological index for simulating the land surface eco-environment. *Journal of Environmental Management*, *326*, 116851. https://doi.org/10.1016/j.jenvman.2022.116851

Williams, C. (2019). Raster Formats and Sources. In *The Geographic Information Science & Technology Body of Knowledge (4th Quarter 2019 Edition) , John P. Wilson (Ed.).* https://doi.org/10.22224/gistbok/2019.4.11.

Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, *80*(3), 385–396. https://doi.org/10.1016/S0034-4257(01)00318-2

Xu, H. (2013). A remote sensing urban ecological index and its application. *Acta Ecol. Sin*, *33*(24), 7853–7862.

Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, *2017*(1), 1353691. https://doi.org/10.1155/2017/1353691

Yagoub, M. M., Tesfaldet, Y. T., Elmubarak, M. G., & Al Hosani, N. (2022). Extraction of Urban Quality of Life Indicators Using Remote Sensing and Machine Learning: The Case of Al Ain City, United Arab Emirates (UAE). *ISPRS International Journal of Geo-Information*, *11*(9), Article 9. https://doi.org/10.3390/ijgi11090458

Zabalza, J., Ren, J., Yang, M., Zhang, Y., Wang, J., Marshall, S., & Han, J. (2014). Novel Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *93*, 112–122. https://doi.org/10.1016/j.isprsjprs.2014.04.006

Zhang, T., Yang, R., Yang, Y., Li, L., & Chen, L. (2021). Assessing the Urban Eco-Environmental Quality by the Remote-Sensing Ecological Index: Application to Tianjin, North China. *ISPRS International Journal of Geo-Information*, *10*(7), Article 7. https://doi.org/10.3390/ijgi10070475

Zhang, X., Jia, W., Lu, S., & He, J. (2024). Ecological assessment and driver analysis of high vegetation cover areas based on new remote sensing index. *Ecological Informatics*, *82*, 102786. https://doi.org/10.1016/j.ecoinf.2024.102786

Zheng, Y., Tang, L., & Wang, H. (2021). An improved approach for monitoring urban built-up areas by combining NPP-VIIRS nighttime light, NDVI, NDWI, and NDBI. *Journal of Cleaner Production*, *328*, 129488. https://doi.org/10.1016/j.jclepro.2021.129488

Zhu, D., Chen, T., Wang, Z., & Niu, R. (2021). Detecting ecological spatial-temporal changes by Remote Sensing Ecological Index with local adaptability. *Journal of Environmental Management*, *299*, 113655. https://doi.org/10.1016/j.jenvman.2021.113655

Appendix

Table A-1 Results of all major PCs and their variance (Tallinn).

PCs	Eigenvalues	Prop. of Variance	Cumulative Prop. of Variance
PC1	2.535	0.507	0.507
PC2	2.079	0.416	0.923
PC3	0.235	0.047	0.970
PC4	0.130	0.026	0.996
PC5	0.021	0.004	1.000

Table A-2 Results of all major PCs and their variance (Tartu).

PCs	Eigenvalues	Prop. of Variance	Cumulative Prop. of Variance
PC1	3.952	0.790	0.790
PC2	0.567	0.113	0.904
PC3	0.393	0.079	0.982
PC4	0.074	0.015	0.997
PC5	0.015	0.003	1.000

Table A-3 Results of all major PCs and their variance (Porijõgi Catchment).

PCs	Eigenvalues	Prop. of Variance	Cumulative Prop. of Variance
PC1	3.322	0.664	0.664
PC2	0.841	0.168	0.833
PC3	0.659	0.132	0.964
PC4	0.168	0.034	0.998
PC5	0.011	0.002	1.000

Table A-4 Results of all major PCs and their variance (Ida-Viru Mining Region).

PCs	Eigenvalues	Prop. of Variance	Cumulative Prop. of Variance
PC1	3.098	0.620	0.620
PC2	1.280	0.256	0.876
PC3	0.410	0.082	0.958
PC4	0.162	0.032	0.990
PC5	0.050	0.010	1.000

Table A-5 PC1 and PC2 Loadings of Indices (Tallinn).

Index	PC1_loading	PC2_loading
NDVI	0.507	0.392
NDWI	-0.441	-0.474
NDMI	0.501	-0.344
BSI	-0.534	0.313
LST	-0.108	0.636

Table A-6 PC1 and PC2 Loadings of Indices (Tartu).

Index	PC1_loading	PC2_loading
NDVI	0.476	0.306
NDWI	-0.429	-0.399
NDMI	0.475	-0.020
BSI	-0.479	-0.020
LST	-0.367	0.864

Table A-7 PC1 and PC2 Loadings of Indices (Porijõgi Catchment).

Index	PC1_loading	PC2_loading
NDVI	0.499	0.065
NDWI	-0.370	-0.135
NDMI	0.520	0.267
BSI	-0.523	-0.067
LST	-0.266	0.950

Table A-8 PC1 and PC2 Loadings of Indices (Ida-Viru Mining Region).

Index	PC1_loading	PC2_loading
NDVI	0.469	0.451
NDWI	-0.402	-0.603
NDMI	0.478	-0.195
BSI	-0.502	0.312
LST	-0.373	0.546

Table A-9 Threshold values or Index in Tallinn

Index	Threshold value	Rank	Category	Weight
NDVI	> 0.680	I	Very High	0.242

	0.243< X<= 0.680	II	High	
	-0.194< X<= 0.243	III	Moderate	
	<= -0.194	IV	Low	
NDWI	> 0.236	I	Very High	0.211
	-0.144< X<= 0.236	II	High	
	-0.524< X<= -0.144	III	Moderate	
	<= -0.524	IV	Low	
NDMI	> 0.513	I	Very High	0.240
	0.323< X<= 0.513	II	High	
	0.133< X<= 0.323	III	Moderate	
	<= 0.133	IV	Low	
BSI	< = -0.438	I	Very High	0.255
	-0.438< X <= -0.268	II	High	
	-0.268<9 X<= -0.097	III	Moderate	
	> 0.097	IV	Low	
LST	< = 21.475	I	Very High	0.052
	21.475< X <= 27.424	II	High	
	27.424< X<=33.372	III	Moderate	
	> 33.372	IV	Low	

Table A-10 Threshold values of Index in Tartu.

Index	Threshold value	Rank	Category	Weight
NDVI	> 0.732	I	Very High	0.214
	0.469< X<= 0.732	II	High	
	0.205< X<= 0.469	III	Moderate	
	<= 0.205	IV	Low	
NDWI	> -0.225	I	Very High	0.193
	-0.422< X<= -0.225	II	High	
	-0.620< X<= -0.422	III	Moderate	
	<= -0.620	IV	Low	
NDMI	> 0.278	I	Very High	0.213
	0.107< X<= 0.278	II	High	
	-0.063< X<= 0.107	III	Moderate	
	<= -0.063	IV	Low	
BSI	< = -0.233	I	Very High	0.215
	-0.233< X <= -0.071	II	High	
	-0.071< X<= 0.090	III	Moderate	
	> 0.090	IV	Low	
LST	< = 34.550	I	Very High	0.165
	31.550< X <= 35.692	II	High	
	35.692< X<=39.834	III	Moderate	
	> 39.834	IV	Low	

Table A-11 Threshold values of Index in Porijõgi Catchment.

Index	Threshold value	Rank	Category	Weight
NDVI	> 0.837	I	Very High	0.229
	0.686< X<= 0.837	II	High	
	0.536< X<= 0.686	III	Moderate	
	<= 0.536	IV	Low	
NDWI	> -0.424	I	Very High	0.170
	-0.571< X<= -0.424	II	High	
	-0.718< X<= -0.571	III	Moderate	
	<= -0.718	IV	Low	
NDMI	> 0.379	I	Very High	0.239
	0.242< X<= 0.379	II	High	
	0.104< X<= 0.242	III	Moderate	
	<= 0.104	IV	Low	
BSI	< = -0.340	I	Very High	0.240
	-0.340< X <= -0.197	II	High	
	-0.197< X<= -0.055	III	Moderate	
	> -0.055	IV	Low	
LST	< = 27.871	I	Very High	0.122
	30.050< X <= 27.871	II	High	
	32.288< X<=30.050	III	Moderate	
	> 32.288	IV	Low	

Table A-12 Threshold values of Index in Ida-Viru Mining Region.

Index	Threshold value	Rank	Category	Weight
NDVI	> 0.835	I	Very High	0.203
	0.563< X<= 0.835	II	High	
	0.291< X<= 0.563	III	Moderate	
	<=0.291	IV	Low	
NDWI	> -0.172	I	Very High	0.174
	-0.418< X<= -0.172	II	High	
	-0.664< X<= -0.418	III	Moderate	
	<= -0.664	IV	Low	
NDMI	> 0.509	I	Very High	0.207
	0.324< X<= 0.509	II	High	
	0.139< X<= 0.324	III	Moderate	
	<= 0.139	IV	Low	
BSI	< = -0.341	I	Very High	0.256
	-0.341< X <= -0.166	II	High	
	-0.166< X<= -0.009	III	Moderate	
	> 0.009	IV	Low	
LST	< = 30.116	I	Very High	0.161
	30.116< X <=34.038	II	High	
	34.038< X<=37.960	III	Moderate	
	> 37.960	IV	Low	

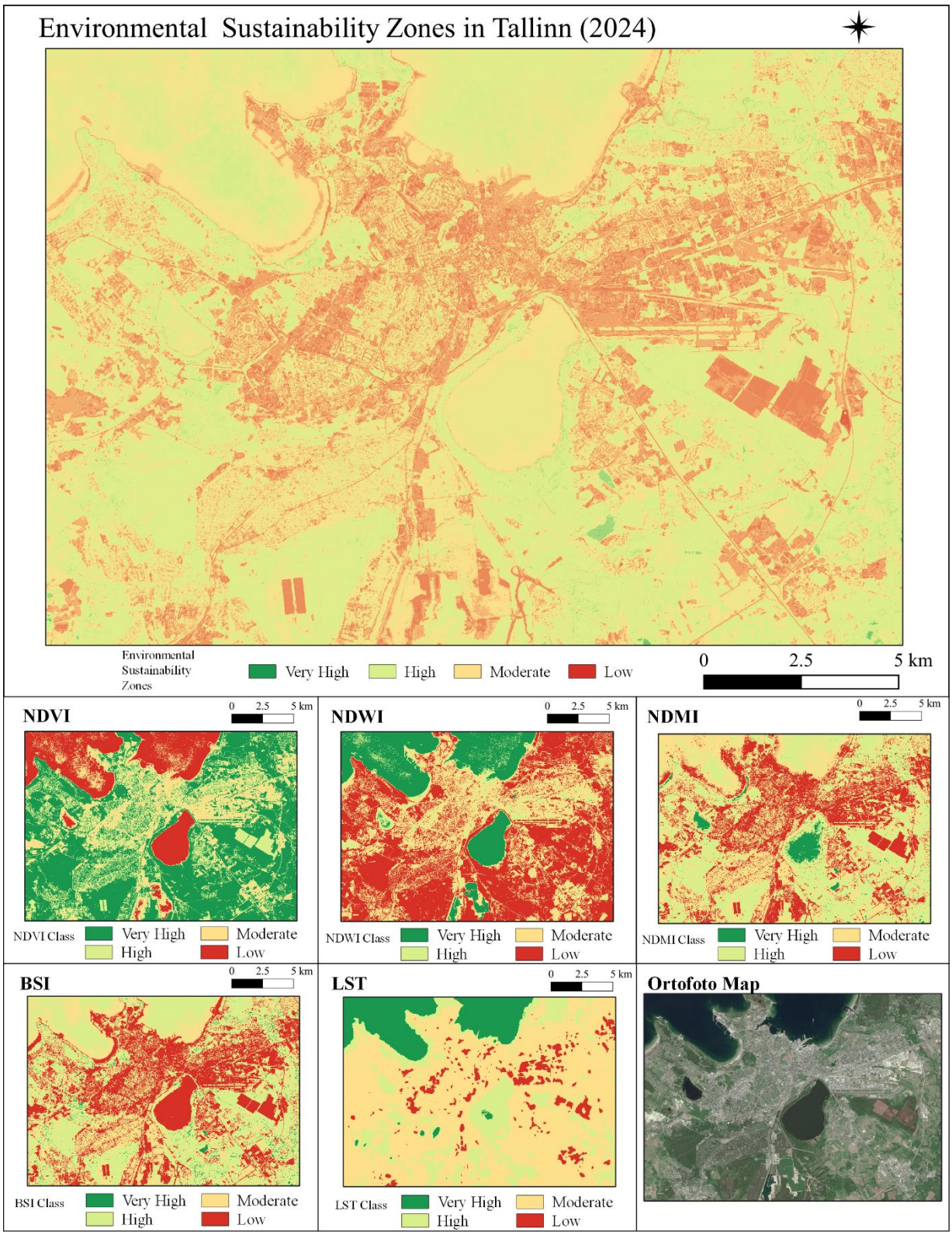


Figure A-1 The environmental sustainability zones map in Tallinn (2024).

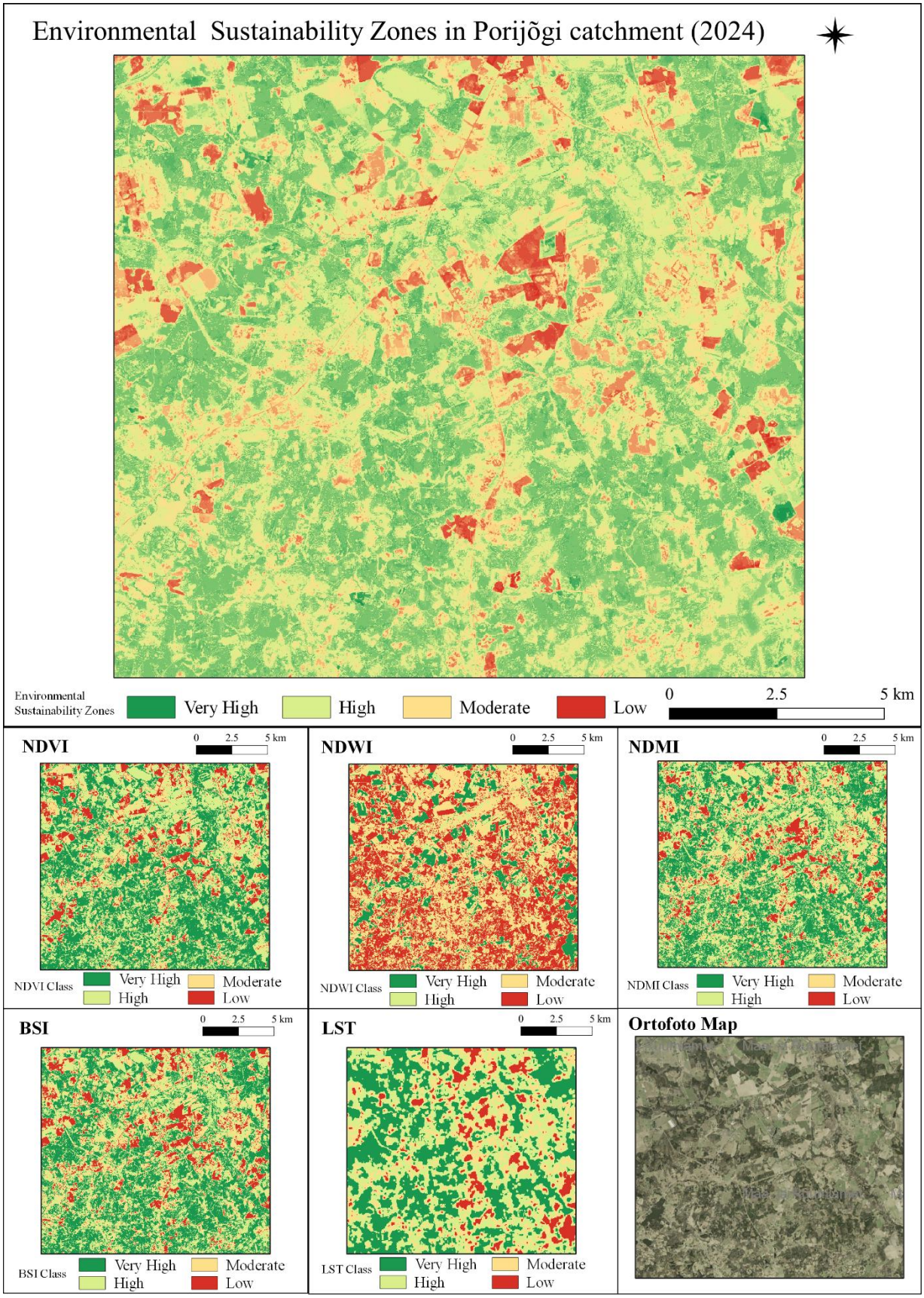


Figure A-2 The environmental sustainability zones map in Porijõgi Catchment (2024).

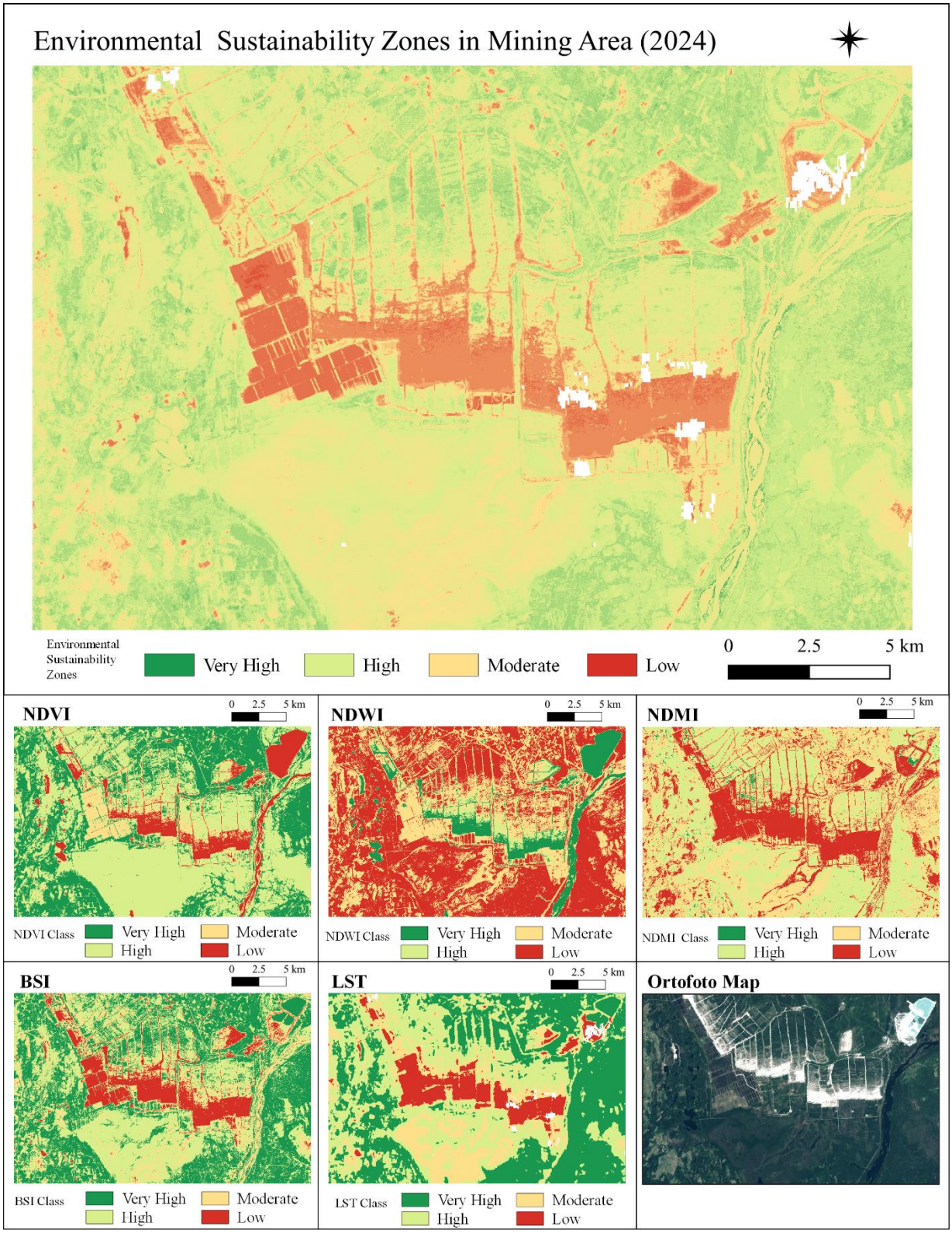


Figure A-3 The environmental sustainability zones map in Ida-Viru Mining Region (2024).

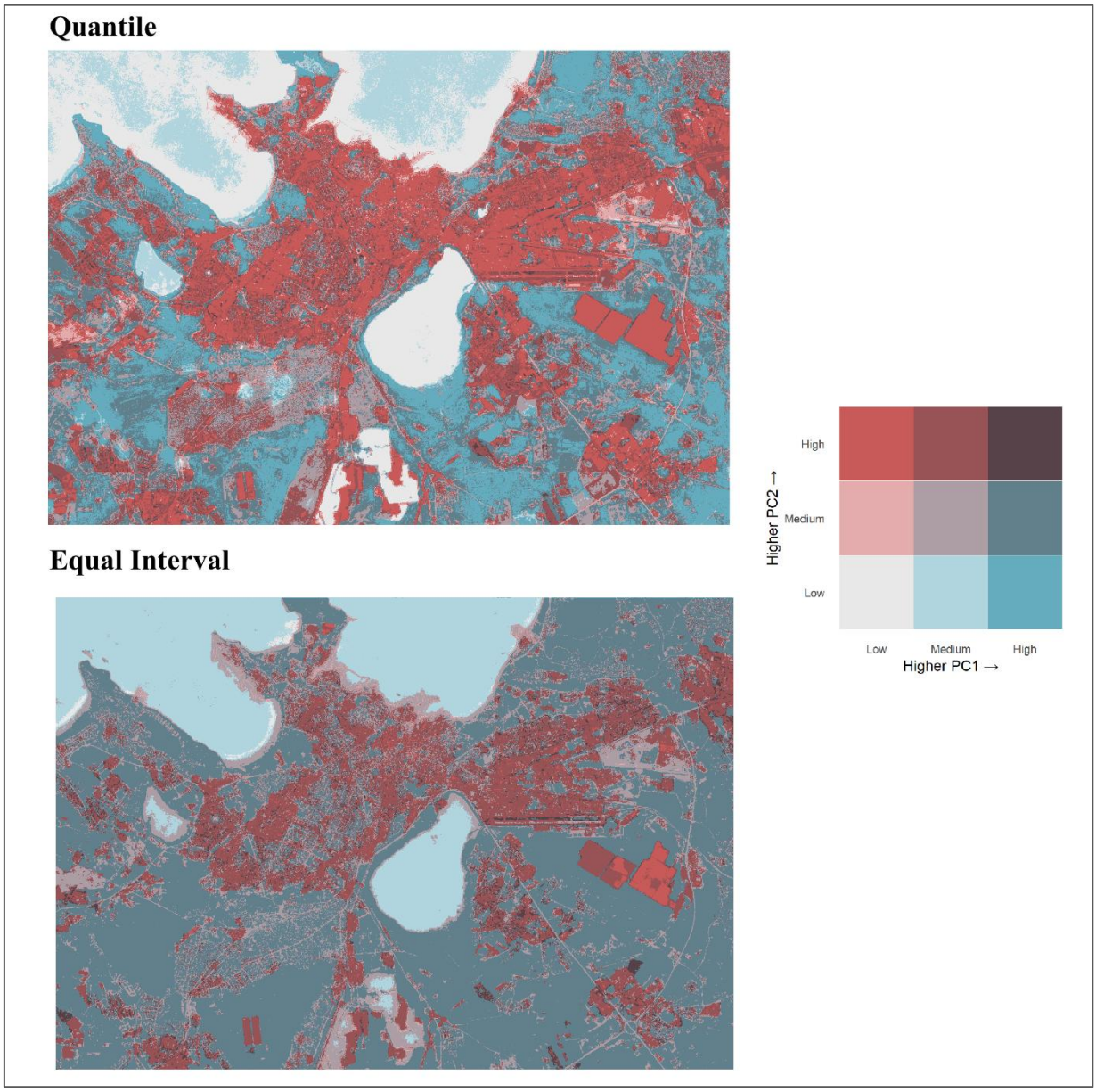


Figure A-4 Standard bivariate raster maps of Tallinn.

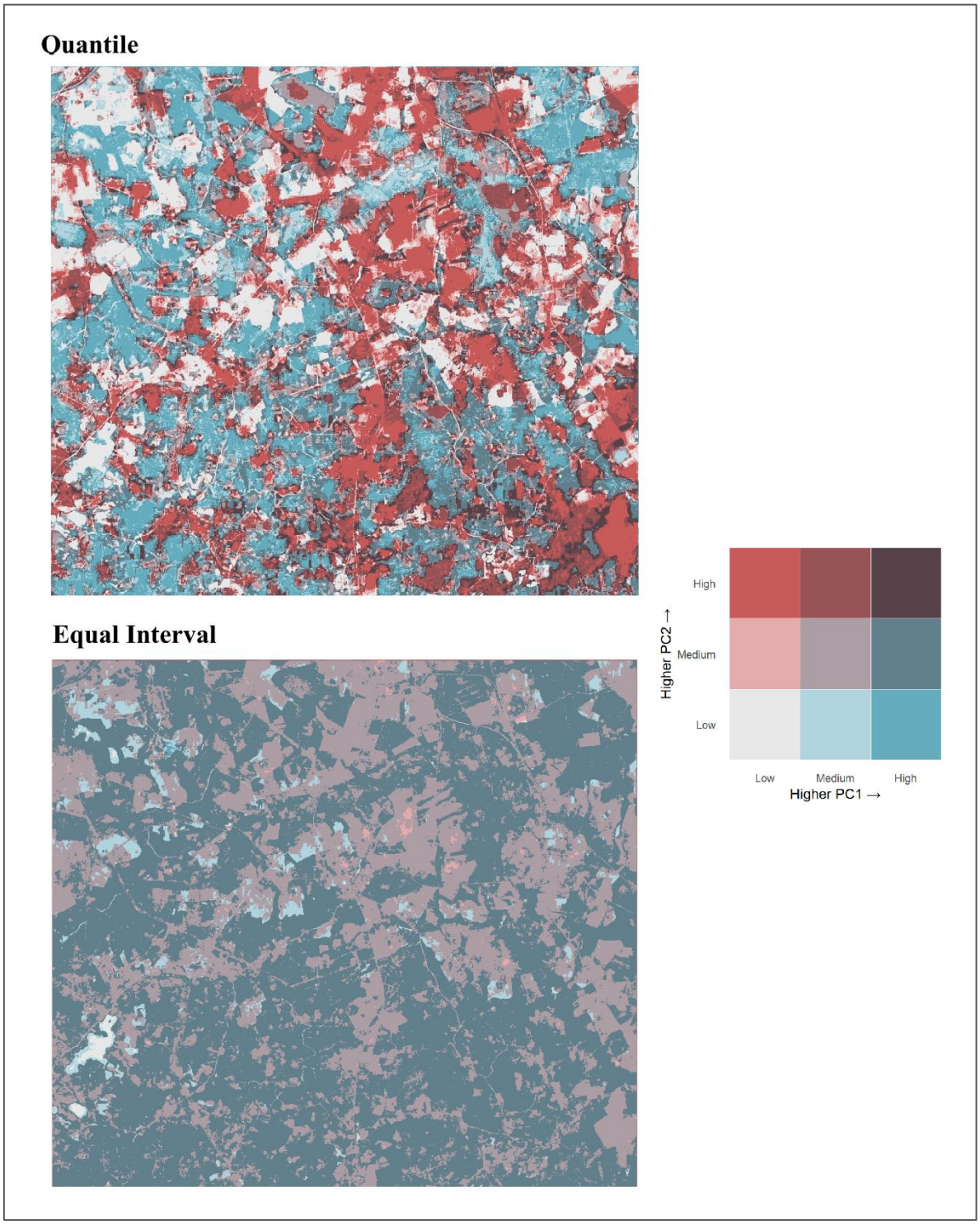


Figure A-5 Standard bivariate raster maps of Porijõgi Catchment.

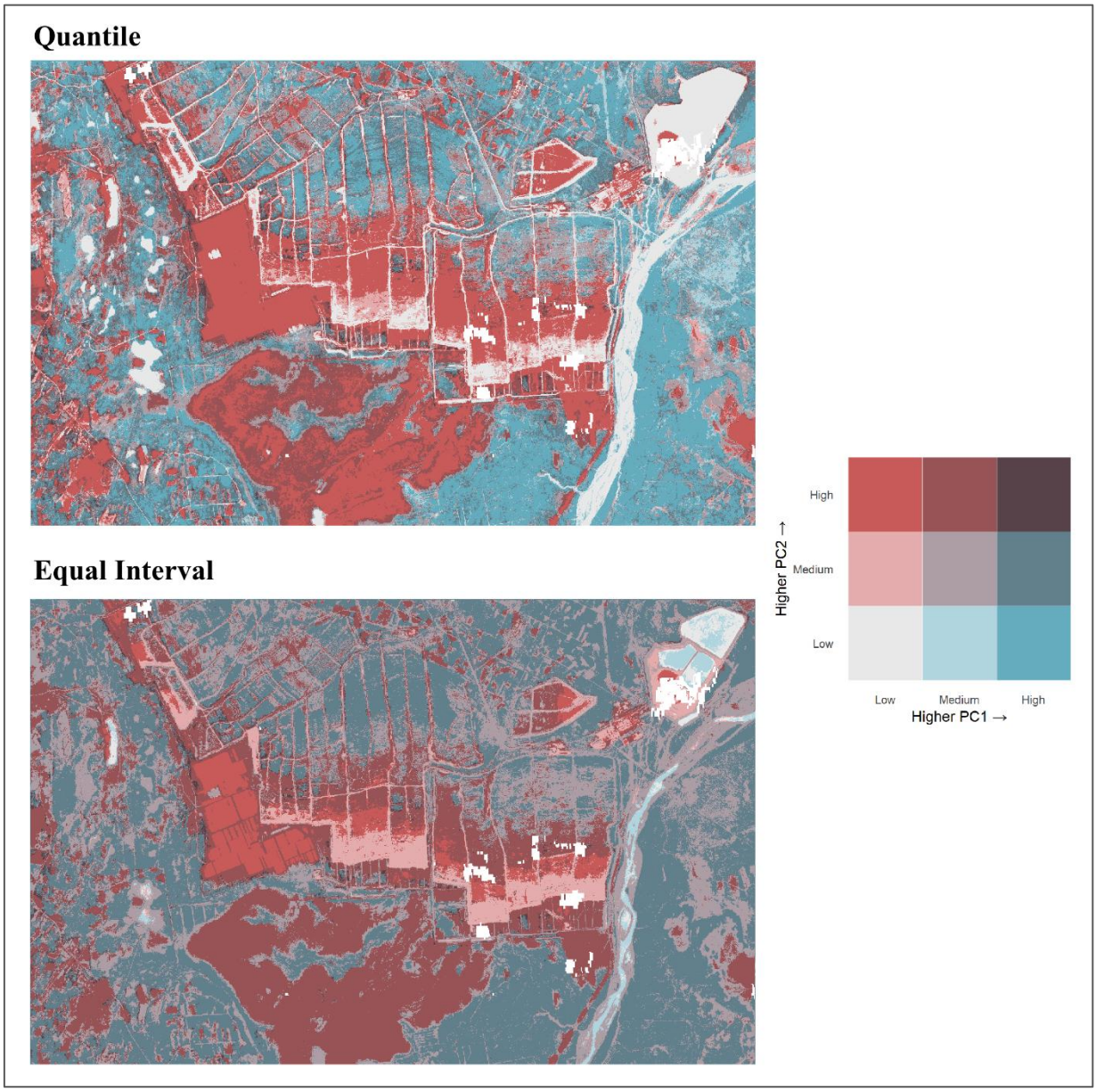


Figure A-6 Standard bivariate raster maps of Ida-Viru Mining Region.

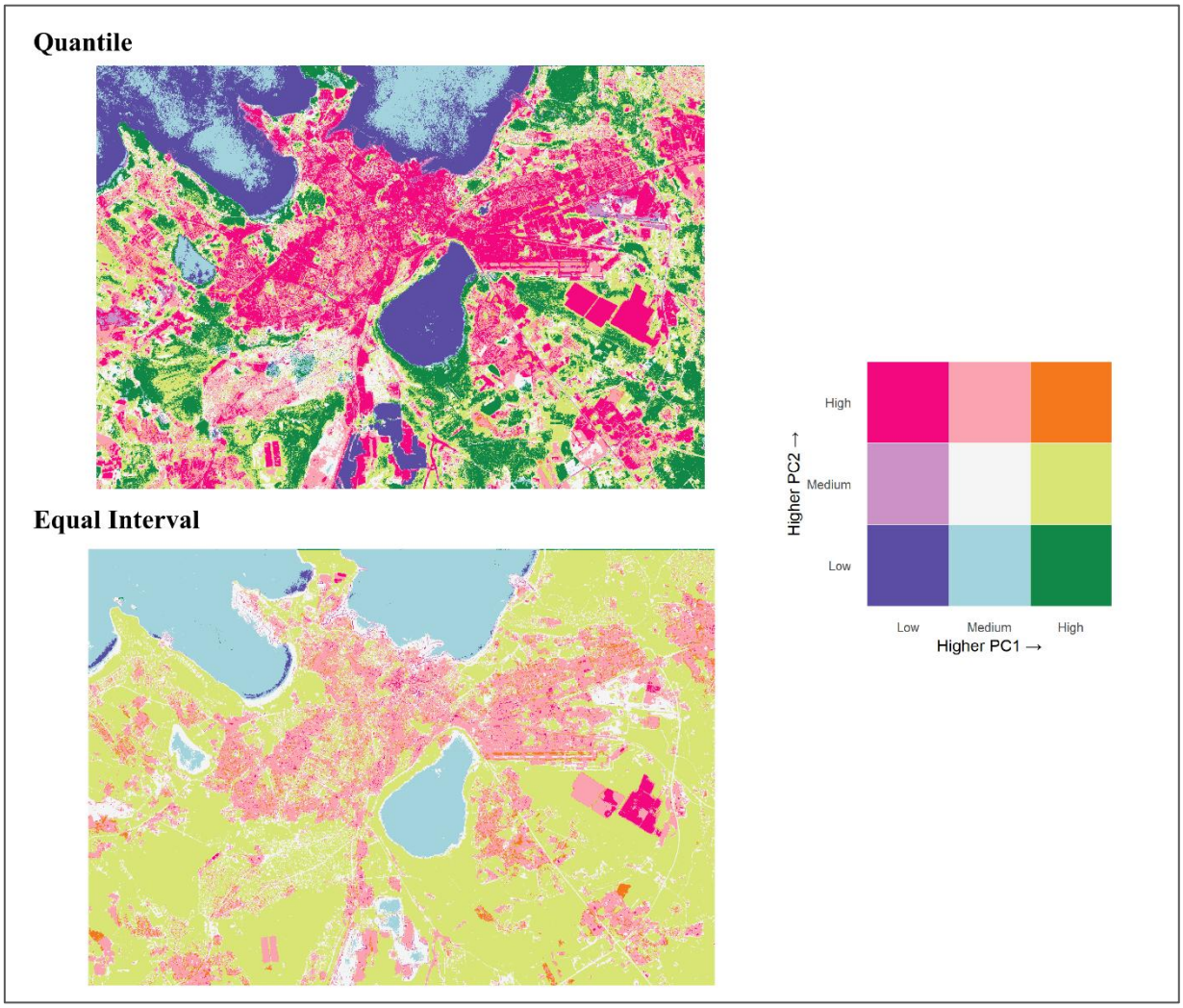


Figure A-7 Corner model applied to Tallinn.

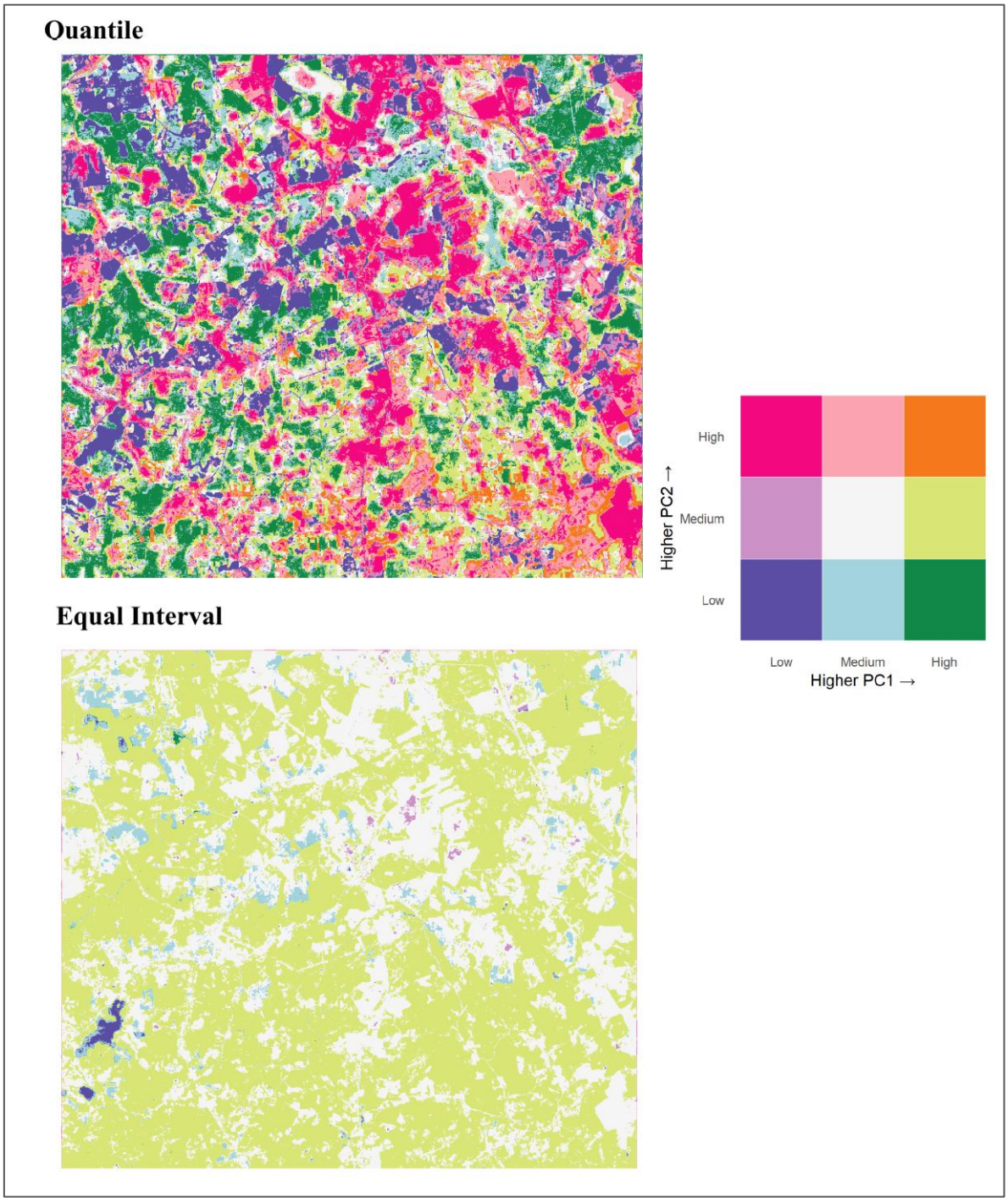


Figure A-8 Corner model applied to Porijõgi Catchment.

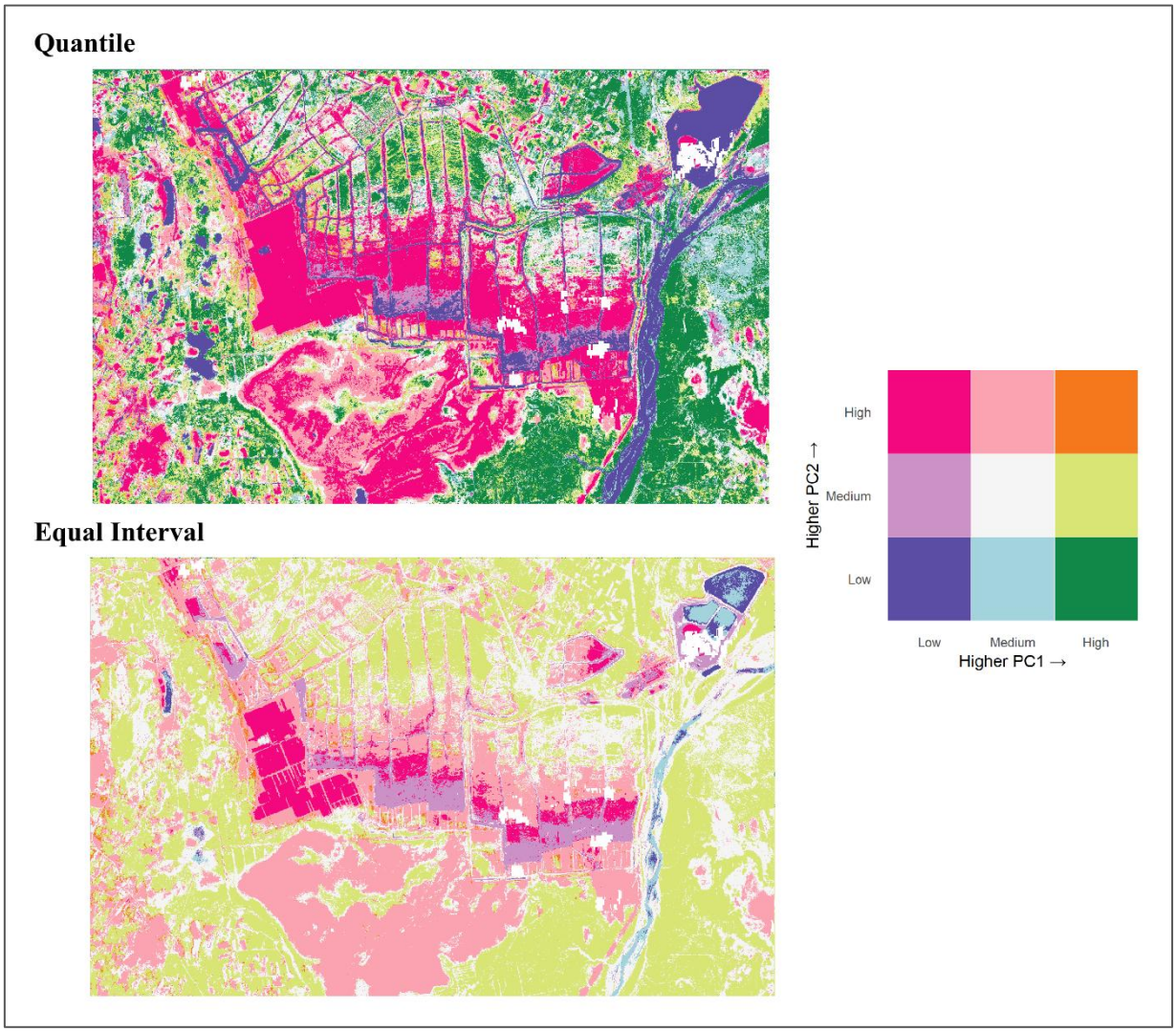


Figure A-9 Corner model applied to Ida-Viru Mining Region.

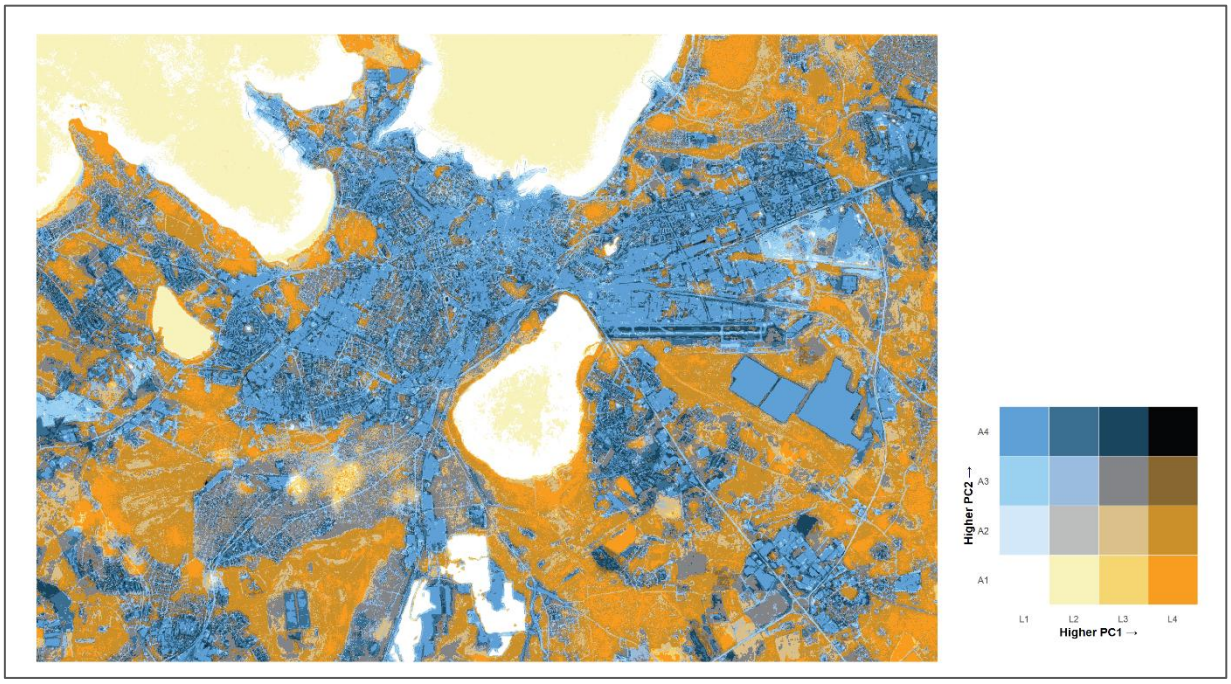


Figure A-10 Diagonal Model applied to Tallinn.

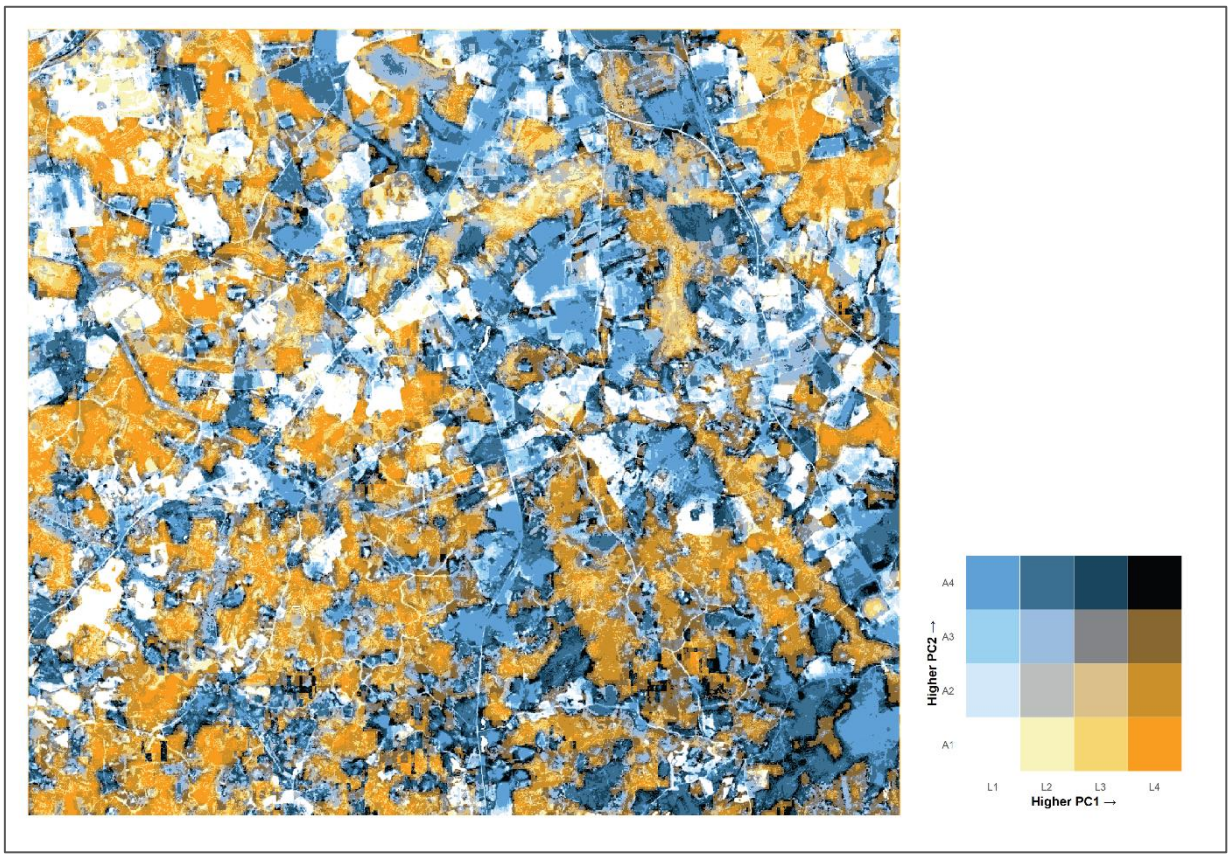


Figure A-11 Diagonal Model applied to Porijõgi Catchment.

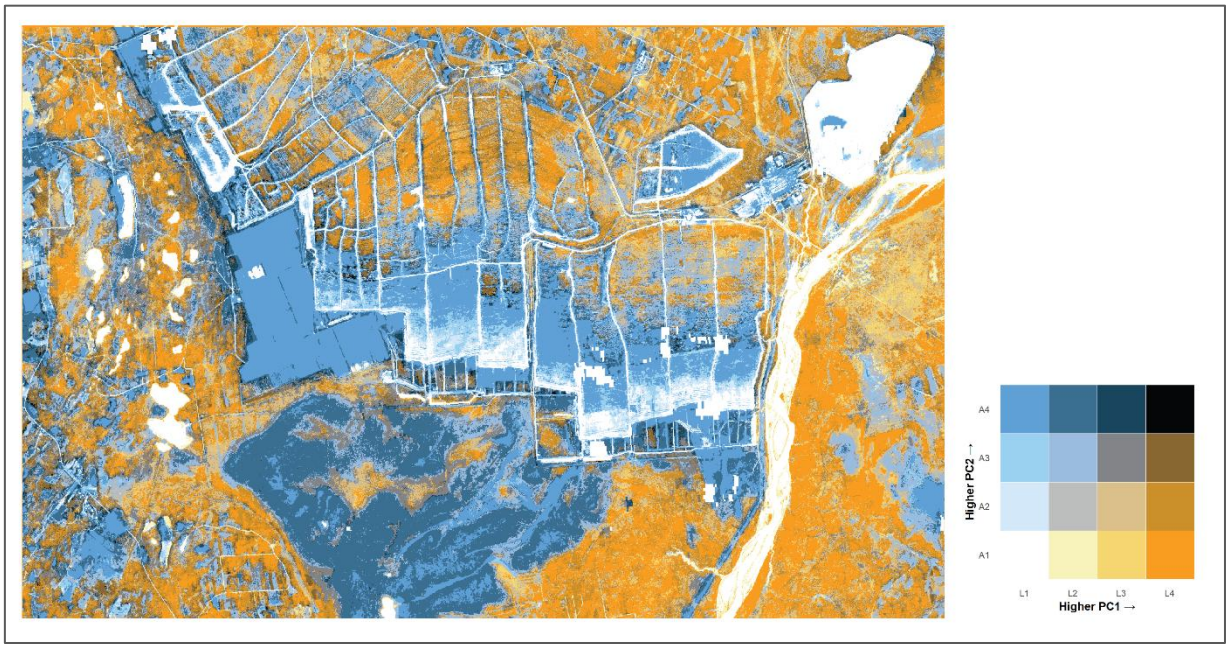


Figure A-12 Diagonal Model applied to Ida-Viru Mining Region.

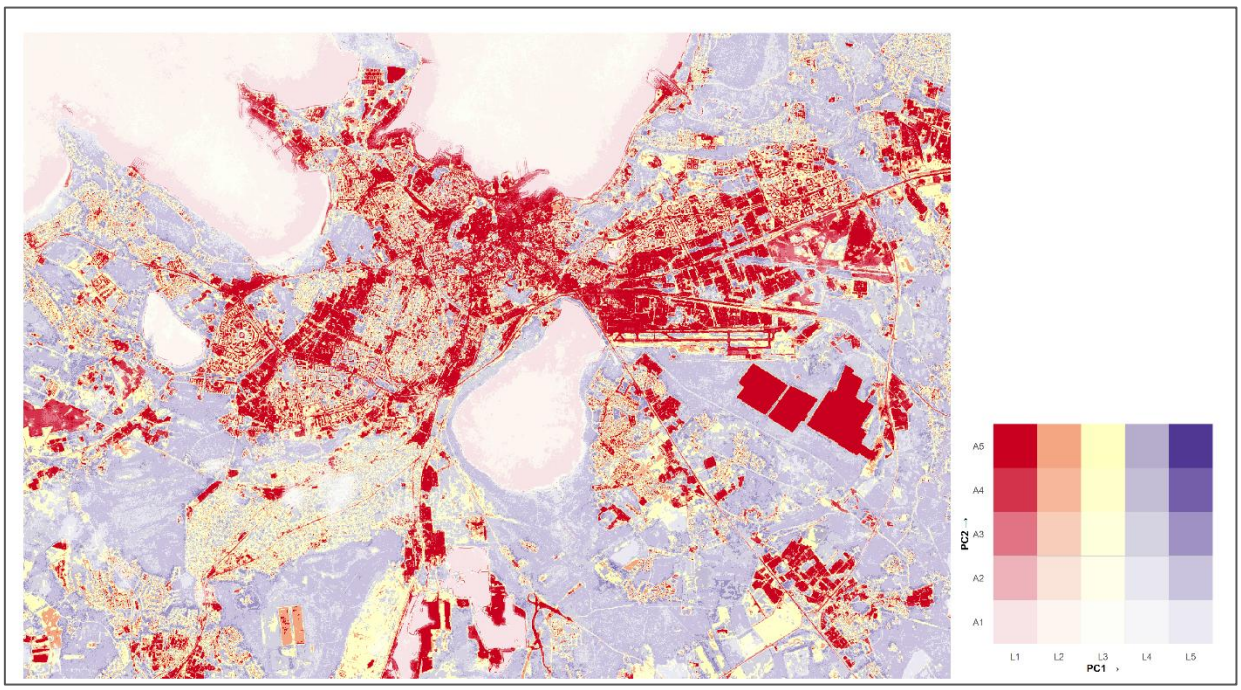


Figure A-13 Value-by-alpha map of Tallinn without background.

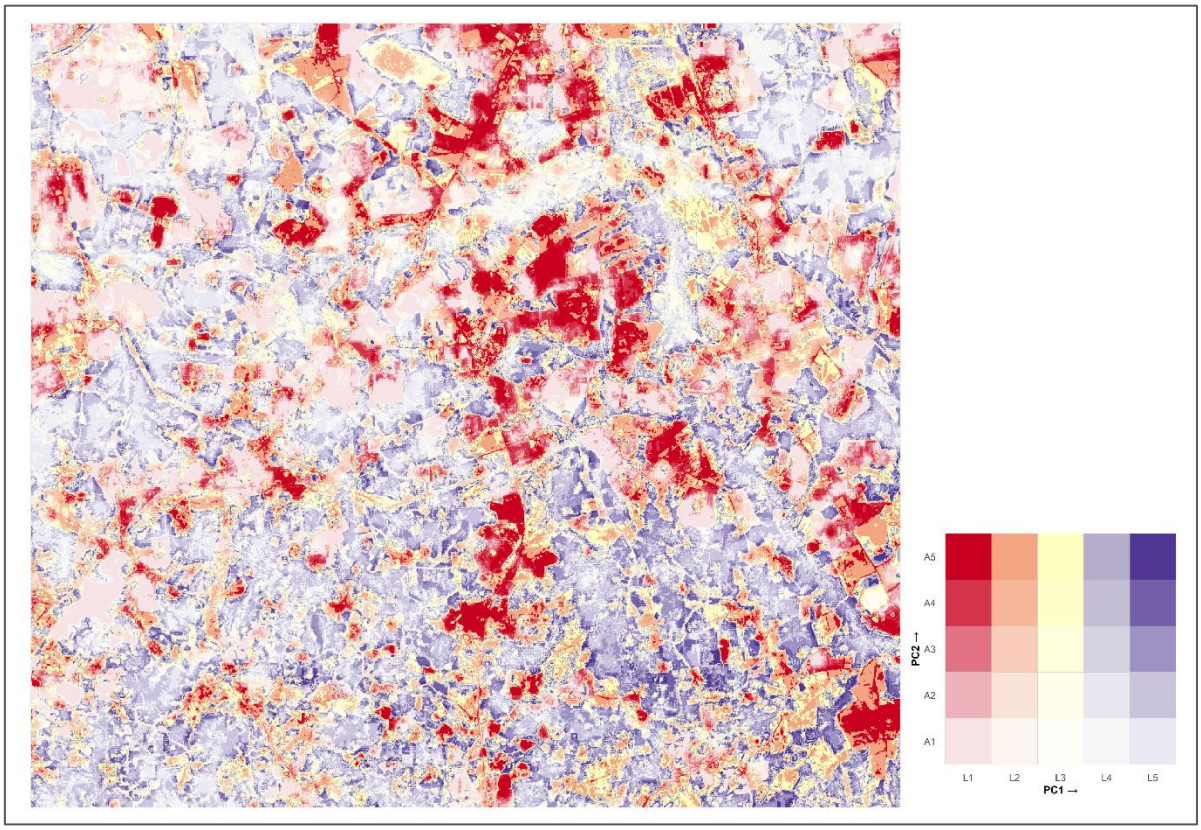


Figure A-14 Value-by-alpha map of Porijõgi Catchment without background.

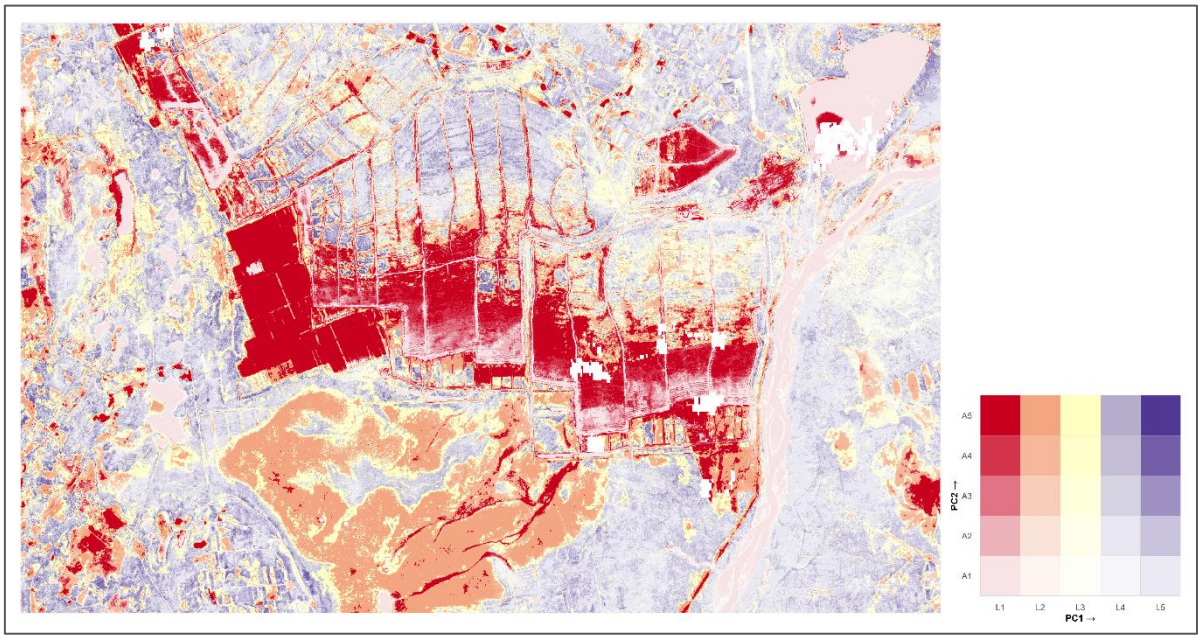


Figure A-15 Value-by-alpha map of Ida-Viru Mining Region without background.

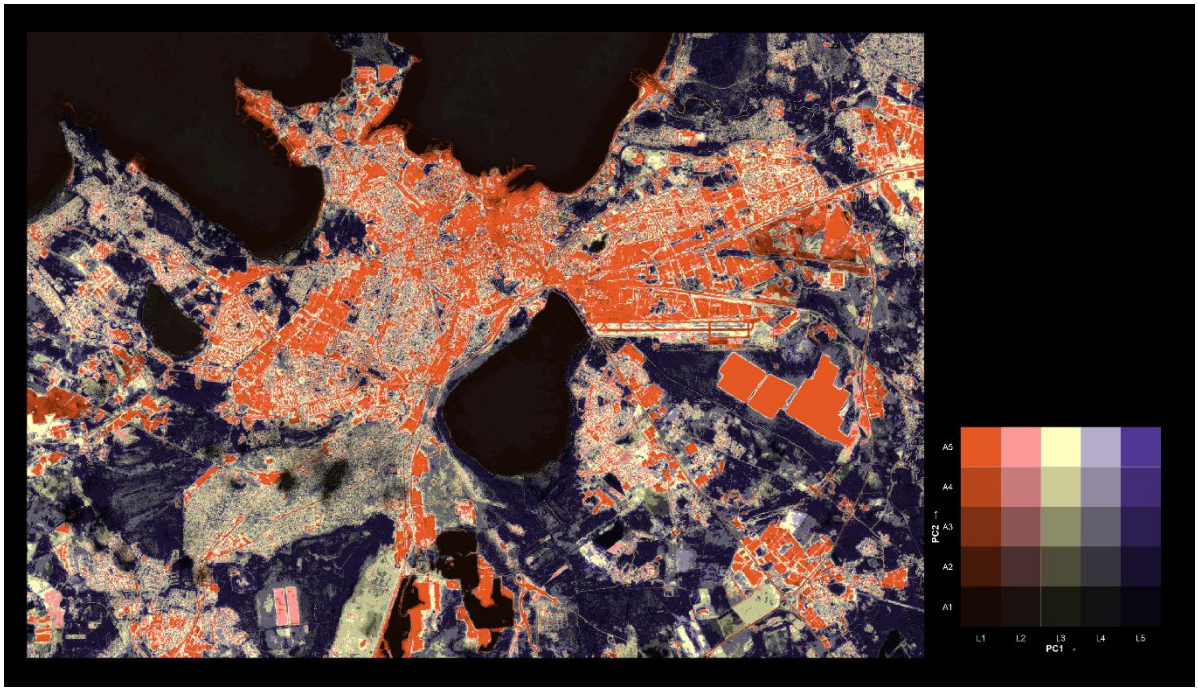


Figure A-16 Black background value-by-alpha spotlight map of Tallinn.

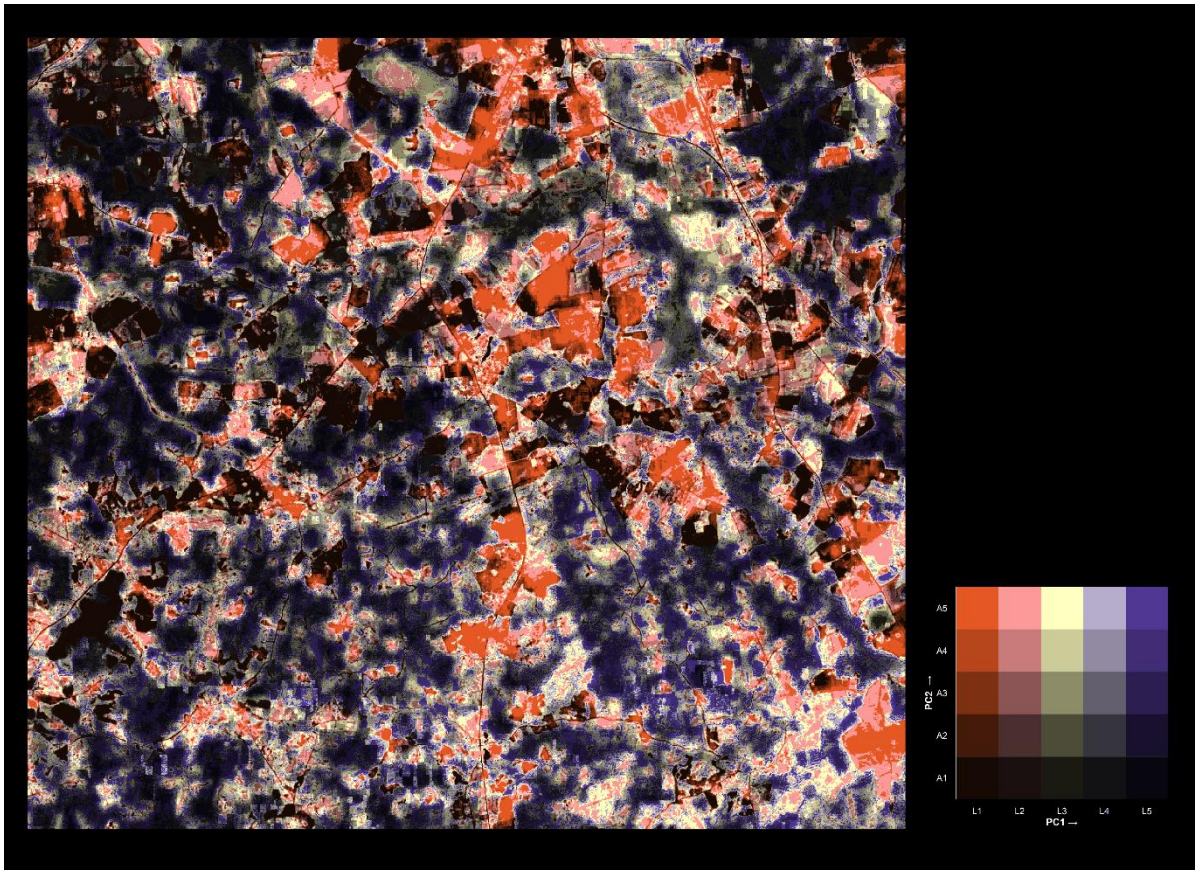


Figure A-17 Black background value-by-alpha spotlight map of Porijõgi Catchment.

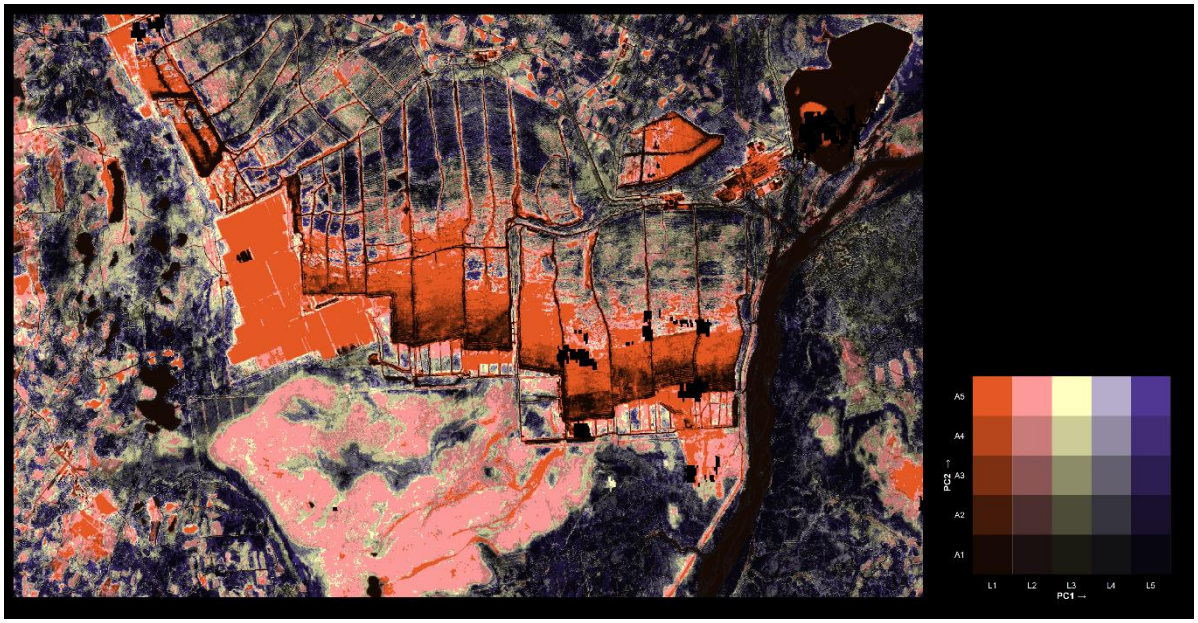


Figure A-18 Black background value-by-alpha spotlight map of Ida-Viru Mining Region.

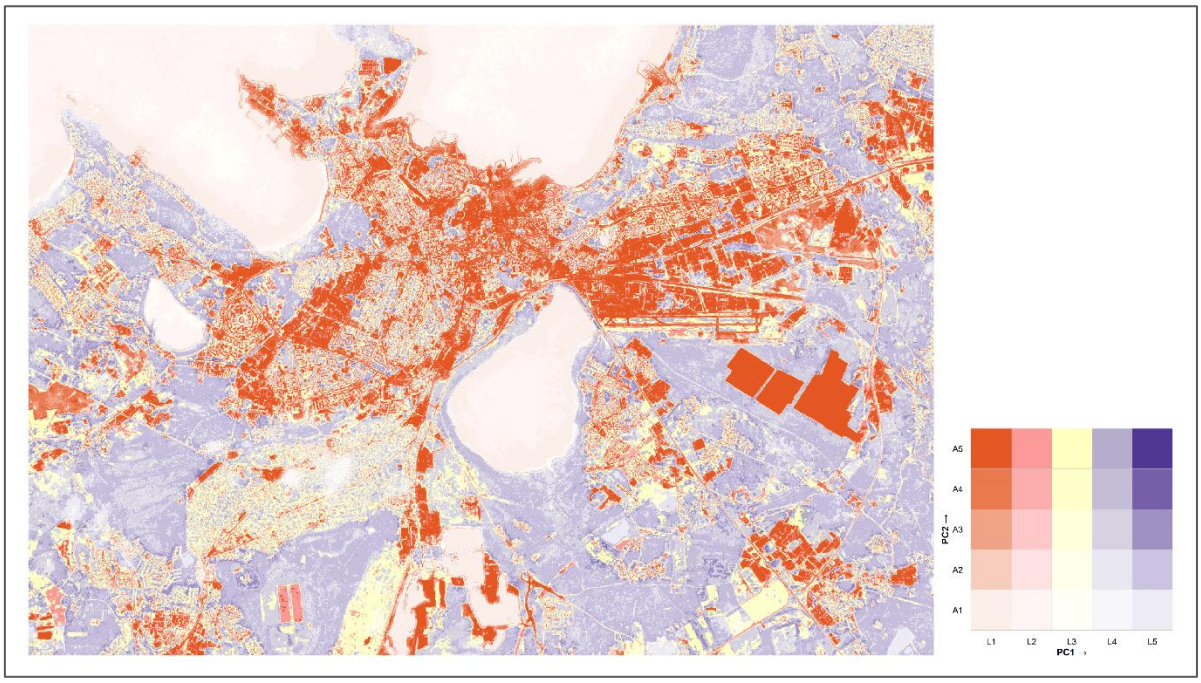


Figure A-19 White background value-by-alpha spotlight map of Tallinn.

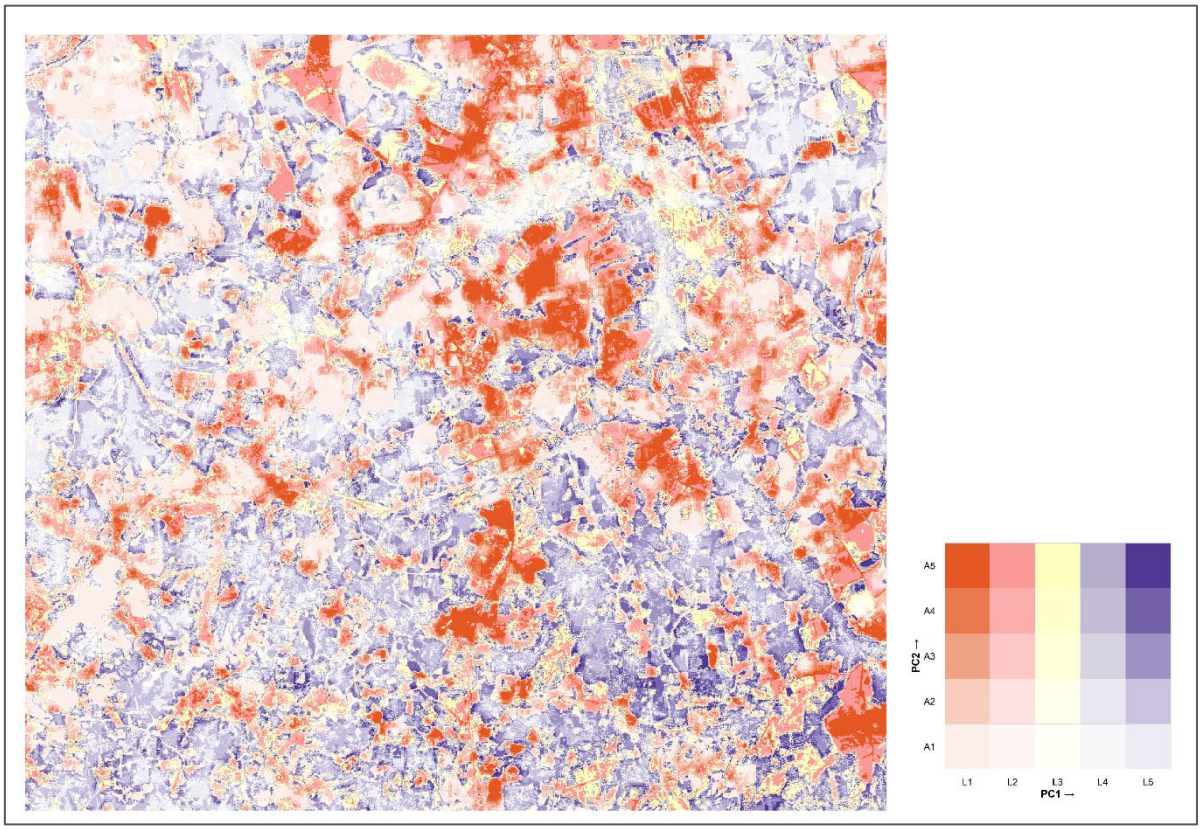


Figure A-20 White background value-by-alpha spotlight map of Porijõgi Catchment.

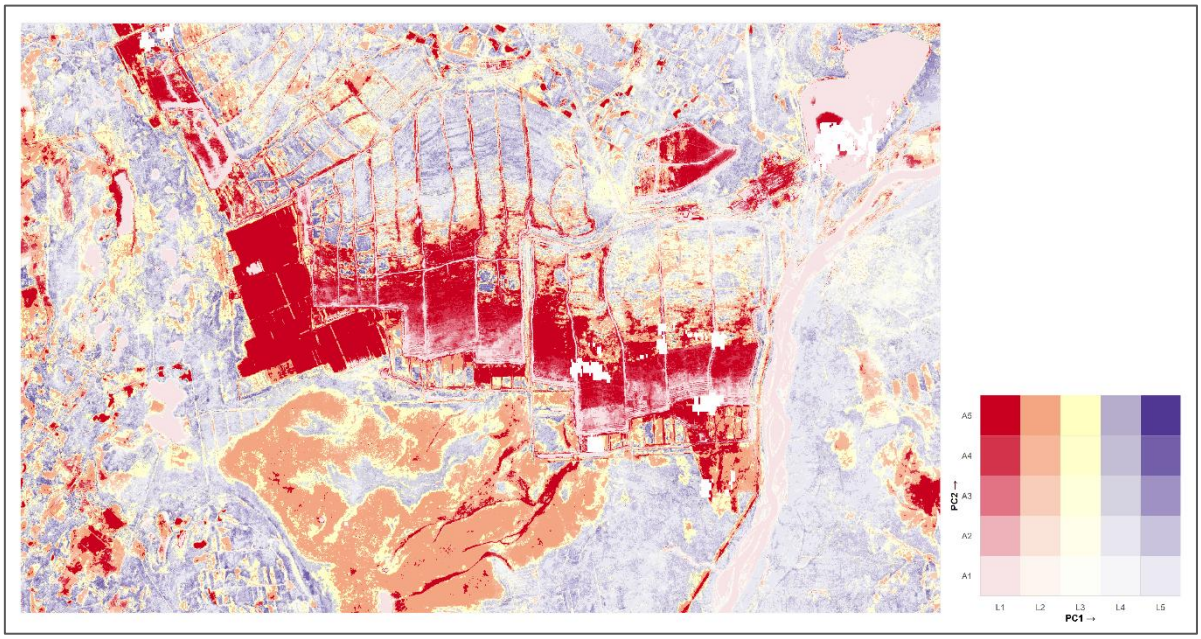


Figure A-21 White background value-by-alpha spotlight map of Ida-Viru Mining Region.

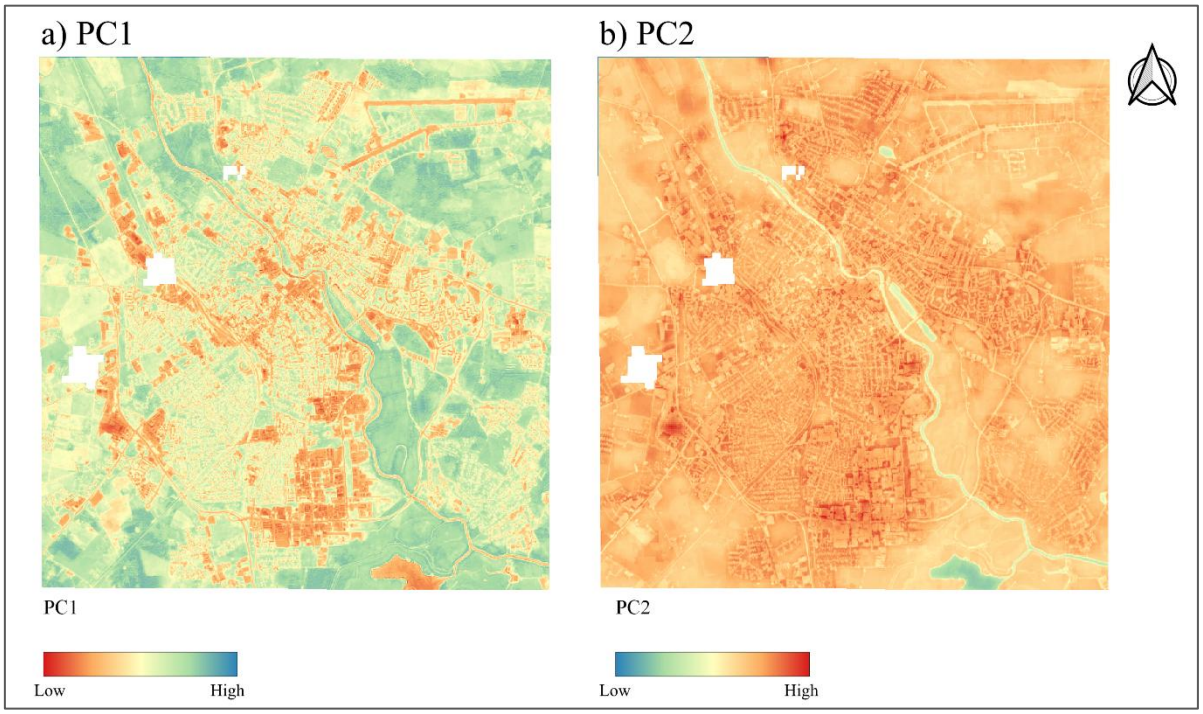


Figure A-22 The PC1 and PC2 of Tartu study area, while the higher PC1 represents better vegetation, higher PC2 represents higher temperature.

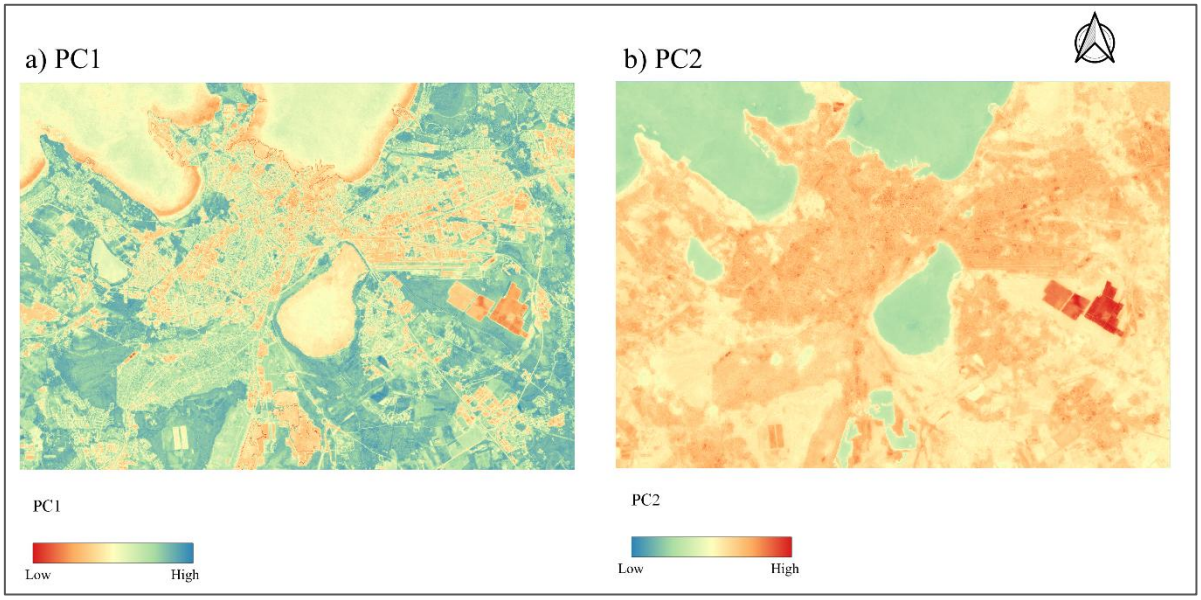


Figure A-23 The PC1 and PC2 of Tallinn study area, while the higher PC1 represents better vegetation, higher PC2 represents higher temperature.

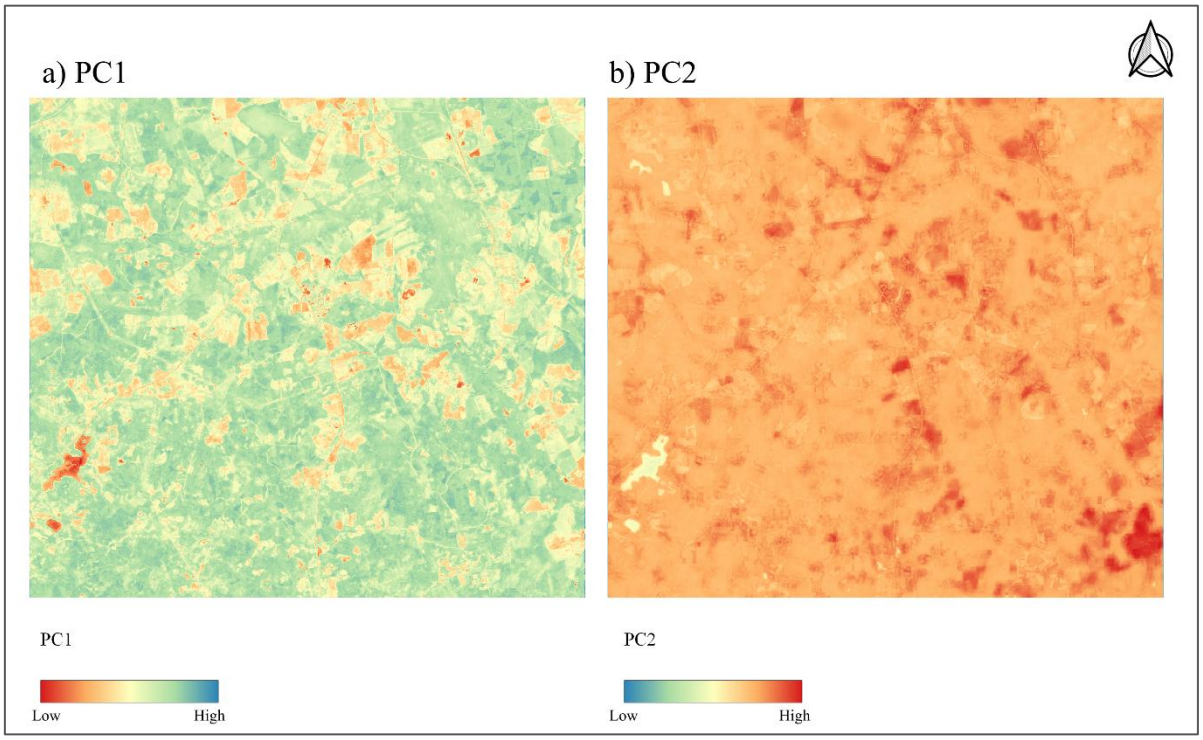


Figure A-24 The PC1 and PC2 of Porijõgi Catchment, while the higher PC1 represents better vegetation, higher PC2 represents higher temperature.

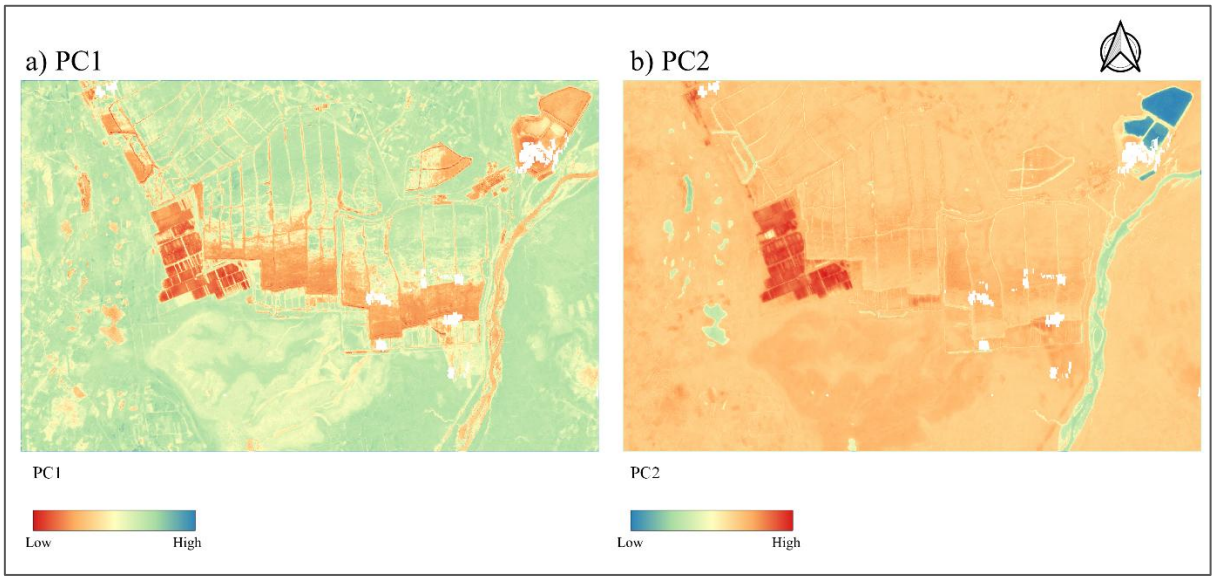


Figure A-25 The PC1 and PC2 of Ida-Viru Mining Region, while the higher PC1 represents better vegetation, higher PC2 represents higher temperature.

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Wenyi Fang ,

(*author's name*)

1. grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

Comparing Raster Visualization Techniques for Environmental Indices: Univariate and Bivariate Mapping in Estonia

(*title of thesis*)

supervised by

Alexander Kmoch

(*supervisor's name*)

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in points 1 and 2.
4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Wenyi Fang

23/05/2025