



ORIGINAL ARTICLE

# Quantifying the Influence of Climate Variables on Vegetation Through Remote Sensing and Multi-dimensional Data Analysis

Hatef Dastour<sup>1</sup> · Quazi K. Hassan<sup>1</sup>

Received: 9 December 2023 / Revised: 27 January 2024 / Accepted: 28 January 2024 / Published online: 16 March 2024  
© King Abdulaziz University and Springer Nature Switzerland AG 2024

## Abstract

The exchange of energy, water, and carbon between the land surface and the atmosphere is critically influenced by vegetation. However, vegetation cover has been changing due to climate variability and human activities, which can affect ecosystems, biodiversity, land management, and human well-being. This study aimed to examine the impact of climate factors on different vegetation types in Alberta, Canada. Remote sensing and climate data from various sites were collected and analyzed using spatial and temporal correlation analysis methods. During the study period of 2001–2022, this study revealed that temporal dynamics between NDVI and climate variables were assessed using NCC analyses on various land cover types. The results revealed a lead of 3.5–4 months of NDVI over the Relative Humidity Average. Additionally, LST-Day demonstrated a lead of two weeks over NDVI, while LST-Night exhibited minimal lag with NDVI, except in regions characterized by sparse vegetation. Furthermore, NDVI displayed a lag of 2–4 weeks behind Precipitation. Land cover dynamics in Alberta from 2001 to 2022 reveal significant trends. Cropland areas, covering nearly 20%, consistently increased with increasing relative humidity, except for deviations in 2001 and 2002. Evergreen Needleleaf forests, constituting around 14.5%, exhibited an upward trend correlated with increased precipitation. Grasslands, comprising 13.8%, showed diminishing coverage despite rising humidity and precipitation. Woody Savannas, accounting for approximately 29%, displayed increased coverage in 2006 but exhibited a declining trend over the study period. These trends highlight the complex interplay between land cover changes and climatic factors in Alberta. The quantification of the influence of climate variables on NDVI revealed the pivotal roles of LST-Day and LST-Night, with average feature importance values of 37.42% and 40.35%, respectively, across all land cover types.

**Keywords** Time series analysis · Vegetation greenness · Cross-correlation analyses · Anomaly analyses · Machine learning · Trend analysis

## 1 Introduction

Climate change has caused substantial changes in vegetation cover, affecting the paramount role of vegetation in regulating the energy, water, and carbon fluxes between the land surface and the atmosphere (Julien and Sobrino 2009; Sekhon et al. 2010; Halofsky et al. 2018). To better understand the ecological processes underlying these

changes, it is essential to quantify the influence of climate on vegetation dynamics (Cui et al. 2018; Bari et al. 2021; Hussien et al. 2023). This involves establishing empirical relationships between climatic variables, such as temperature, precipitation, and humidity, and vegetation attributes, such as composition, distribution, and phenology (Cui et al. 2018; Bari et al. 2021; Hussien et al. 2023). By doing so, patterns can be discerned, changes can be forecasted, and deeper insights into the mechanisms governing ecosystems can be gained (Halofsky et al. 2018; Prevedello et al. 2019). Moreover, ecosystem responses to future climate scenarios can be predicted by this quantitative approach (Cui et al. 2018; Bari et al. 2021; Hussien et al. 2023). By extrapolating the observed relationships, potential shifts in vegetation patterns can be anticipated, which have practical implications for land management, conservation efforts, and climate

✉ Quazi K. Hassan  
qhassan@ucalgary.ca

Hatef Dastour  
hatef.dastour@ucalgary.ca

<sup>1</sup> Department of Geomatics Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

change mitigation (Halofsky et al. 2018; Prevedello et al. 2019).

Regional variations play a pivotal role in shaping the impact of climate on vegetation. These variations underscore the diverse climatic conditions, topographical features, and ecological contexts prevalent across different geographic areas (Zhang et al. 2023). Firstly, climatic factors, such as temperature, precipitation, and their seasonality, differ significantly from one region to another. For example, vegetation in arid regions may be more sensitive to changes in precipitation, while vegetation in temperate zones may be more influenced by temperature fluctuations (Wu et al. 2023). Therefore, the climatic variables that have the strongest impact on vegetation can vary markedly between regions (Ali 2013; Jia et al. 2020). Secondly, geographical features, such as mountain ranges, bodies of water, and land cover, can create microclimates and modify the effect of broader climatic patterns on vegetation (Gao et al. 2013; Yin et al. 2016; Jia et al. 2020). These local conditions can either amplify or attenuate the climate-vegetation relationship, adding to the regional variability (Ali 2013). Additionally, ecological factors, including species composition, soil characteristics, and disturbance regimes, interact with climate in unique ways in different regions (Zeng et al. 2023; Sun et al. 2020). These ecological nuances can lead to different responses of vegetation to similar climatic conditions (Prevedello et al. 2019).

Understanding the influence of climate on vegetation in Alberta, Canada is crucial because of the region's unique climatic conditions, topography, and ecological settings (Estevo et al. 2022). This understanding is vital for predicting and addressing the impacts of climate change on ecosystems. Research in this area provides valuable insights into how plant communities adapt to changing climate, which can inform conservation, land management, and sustainable development initiatives (Bellard et al. 2012; Grimm et al. 2013). This knowledge can support evidence-based policies and practices that aim to preserve biodiversity, reduce climate change effects, and enhance ecosystem resilience (Downing and Pettapiece 2006; Jiang et al. 2016; Barber et al. 2015; Baldwin et al. 2020).

Several studies have analyzed the relationship between the Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), and climate data (Hassan and Bourque 2010; Chen et al. 2020; Zeren Cetin et al. 2023). Zhao et al. (2011) examined vegetation growth over 22 years using satellite data, focusing on how it responds to climate change in different regions. They found that approximately 30% of vegetated areas experienced a 0.7% annual increase in vegetation during the growing season. This trend was observed across all biomes and bioregions, except for the Sawuer subregion where there was no significant climate change. The study revealed a strong correlation between

vegetation growth and precipitation and evapotranspiration during the growing season, but not with temperature. Interestingly, the researchers also discovered a positive relationship between winter precipitation and vegetation growth. These findings indicate that precipitation plays a crucial role in vegetation growth in this area, including mountain forests and grasslands, which are typically thought to be limited by low temperatures in winter and early spring (Zhao et al. 2011). Liu and Menzel (2016) studied vegetation and climate trends in Rhineland-Palatinate, Germany. They found upward trends in NDVI for all vegetation types, significant increases in temperature, and weak decreases in precipitation. The temperature was the main driver of vegetation variability, while NDVI showed a negative correlation with precipitation. Strong negative correlations between NDVI and precipitation were observed at the 2-month and 8-month scales. Positive correlations between NDVI and temperature were strongest at the 8- to 16-month scales (Liu and Menzel 2016). Al Balasmeh and Karmaker (2019) examined vegetation dynamics in India across four distinct topographical and climatic conditions. It utilized NDVI data from the SPOT satellite's vegetation sensors, as well as temperature and precipitation data from the Asian Precipitation-Highly Resolved Observational Data Integration Towards Evaluation project. The study discovered that vegetation growth initiates when the average temperature reaches or exceeds 10 °C. Further increases in temperature have minimal impact on vegetation growth. Additionally, the study observed a positive relationship between monthly NDVI and precipitation, depending on the forest type and local climate. However, excessive rainfall led to a decline in vegetation growth. The NDVI data displayed a positive trend across all four sites. In the northern region, temperature showed a positive trend, while precipitation exhibited a negative trend. Conversely, in the eastern and western regions, the temperature had a negative trend, while precipitation showed a positive trend (Al Balasmeh and Karmaker 2019). Guo et al. (2020) conducted a study to investigate the influence of climate change on vegetation in Hainan Island, utilizing meteorological observation data and NDVI monitoring data spanning from 2000 to 2018. The findings indicate favorable ecological conditions on the island, with positive vegetation trends observed. The study highlights the significant positive impact of air temperature on NDVI in Hainan Island. Additionally, it reveals that precipitation hampers vegetation growth, particularly during the rainy season, primarily affecting the central areas of the island. These findings offer a scientific foundation for promoting ecologically sustainable development in the region and provide valuable insights into the effects of climate change on vegetation in tropical island environments (Guo et al. 2020). Chen et al. (2021) analyzed a 33-year time series of Landsat EVI data with machine learning to assess vegetation changes in the western Canadian Arctic.

They found that over 68% of the area experienced significant greening, particularly in shrub-dominated regions with moderate slopes and lower elevations. Dastour et al. (2022) employed the Least-Squares Wavelet (LSWAVE) to examine the connection between climate and vegetation time series by assessing coherency and time lag estimation. Their findings indicate that the seasonal patterns of climate and vegetation exhibit coherence with a time delay. In the Athabasca River Basin (ARB), the most significant coherence was observed in the annual cycle, with vegetation and temperature displaying an 84% annual coherency, and vegetation and precipitation exhibiting a 46% coherency. Specifically, the annual cycles of temperature and precipitation preceded those of vegetation by approximately 2 and 3 weeks, respectively.

The purpose of this research was to examine how climate affects various types of land cover in Alberta, Canada using remote sensing and combined climate data. It studied how temperature, humidity, and precipitation could affect vegetation patterns. A novel method was applied to interpolate climate data. The study also explored the vegetation changes during unusual periods and quantified the nonlinear relationship between climate and vegetation, considering the environmental factors mentioned above. The main objective of this research was to enhance the understanding of vegetation growth and health through the analysis of spatiotemporal parameters. The following analyses were conducted:

- Spatial and temporal correlation assessments.
- Identification of trends and anomalies.
- Quantification of the (nonlinear) influence of climate on vegetation.

## 2 Materials and Methods

### 2.1 Study Region

The province of Alberta, Canada, was the focus of this research, with its vegetation and physical features being the main interest. The study area covered 438,063 square kilometers, extending from 55° to 60° north latitude at the border with the Northwest Territories (Stralberg et al. 2018).

### 2.2 Data

The study employed interpolated weather data sourced from the Agriculture and Irrigation Department (Agriculture and Irrigation Department 2023) for Alberta townships. This dataset encompassed the average relative humidity expressed as a percentage (%) and precipitation measured in millimeters (mm).

The vegetation analysis used various datasets from Google Earth Engine (GEE), a cloud-based platform

for geospatial analysis. The datasets included Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI, MOD13A1 (V061), MODIS LST-Day, MOD11A2 (v061), MODIS LST-Night, MOD11A2 (v061), and MODIS Land Cover Type-1 (LULC), MCD12Q1 (V061), all provided by the National Aeronautics and Space Administration (NASA). The NDVI imagery covered 22 years from 2000 to 2021, while the MODIS annual LULC was available from 2001 to 2021. This study used 16-day MODIS NDVI imagery at 500 m spatial resolution, 8-day MODIS LST-Day and LST-Night at 1 km spatial resolution, and the annual MODIS LULC at 500 m resolution (Sulla-Menashe and Friedl 2018). The pixel quality band of the MODIS data product was used to mask out cloudy pixels and ensure accurate results.

## 2.3 Methods

### 2.3.1 Data Preparation

Climate datasets were generated by integrating interpolated weather data for Alberta townships. For each township, climate attributes such as Relative Humidity Average (%) and Precipitation (mm) were organized based on the geographic coordinates of the township center. The procedure involved the following sequential steps:

- Initial Data Preparation:** The process began with generating sets of 60 images, each with daily temporal resolution. These images were derived from the geographic coordinates of the township centers and then converted into GeoTIFF format.
- Imputation of Potential Missing Values:** Sets of sixty two-dimensional matrices were consolidated into a three-dimensional matrix. Before the upscaling process, missing values in the dataset were addressed through the application of 1D and 2D splines (Alba et al. 2015; Usman and Ramdhani 2019). This addressed gaps both in the spatial xy plane and along the temporal axis. The temporal dimension was represented by a 1D array comprising 60 elements, matching the 60 matrices in the dataset.
- Upscaling:** Spline techniques were employed to enhance the resolution of each individual 3D matrix. This meticulous process ensured that all 60 segments within the matrix effectively accommodated data with a spatial resolution of 500 m.
- Final Separation:** This step involved dividing each enhanced 3D matrix into 60 distinct geo-images. Each image was associated with a specific time step and maintained a consistent spatial resolution of 500 m. At this stage, quality assurance control was performed by comparing the original dataset to the output data.

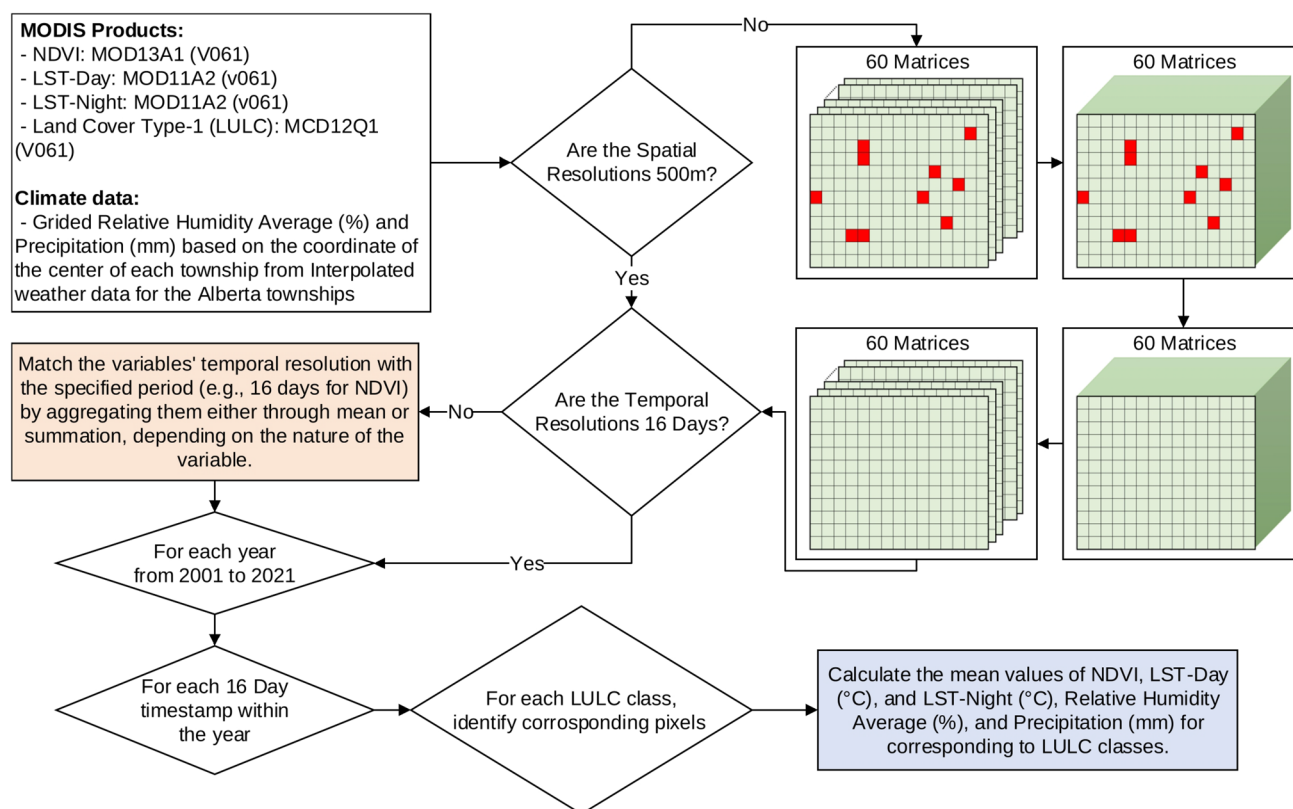
The aggregation of datasets was conducted based on their inherent attributes within the framework of spatiotemporal analyses. Precipitation data was subjected to summation, while the remaining variables were subjected to averaging. This aggregation process was executed to achieve alignment with the 16-day temporal resolution of NDVI. Figure 1 provides a visual representation of the process for generating 16-day datasets that align in both spatial and temporal resolution with NDVI. These datasets encompass LST-Day ( $^{\circ}\text{C}$ ), LST-Night ( $^{\circ}\text{C}$ ), Average Relative Humidity (%), and Precipitation (mm). To access spatiotemporal datasets, one can follow the outlined procedure until reaching the conclusion of the rectangular orange box in Fig. 1.

In the context of land-use land-cover (LULC) class-based analyses, the mean value for each variable was computed individually for each 16-day timestamp and each specific LULC class. To illustrate this process, let's take the example of calculating the Relative Humidity Average (%) for each timestamp and LULC class. This involved utilizing the corresponding LULC pixel distribution from the annual MODIS LULC dataset for the same year. For each timestamp and each LULC class, a value was determined by averaging the

Relative Humidity Average (%) pixels that belonged to the same LULC class. This calculation was carried out consistently for all classes within the given timestamp. This procedure was iterated for all variables, encompassing NDVI, LST-Day ( $^{\circ}\text{C}$ ), LST-Night ( $^{\circ}\text{C}$ ), Relative Humidity Average (%), and Precipitation (mm), utilizing their respective 16-day spatiotemporal datasets and annual LULC classes.

### 2.3.2 Correlation Analyses

The Pearson correlation, a widely employed statistical metric, assesses the strength and direction of a linear relationship between two variables (Lee Rodgers and Nicewander 1988). For spatial analysis, this study computed Pearson correlations between NDVI and the following variables: LST-Day ( $^{\circ}\text{C}$ ), LST-Night ( $^{\circ}\text{C}$ ), Relative Humidity Average (%), and Precipitation (mm). To ensure congruity in spatial and temporal resolution, the methodology detailed in the Sect. 2.3.1 was employed. Then, a pair of time series originating from the two variables (e.g., NDVI and LST-Day) for each pixel were extracted, followed by the calculation of their linear correlation.



**Fig. 1** Workflow for generating hybrid datasets. In this workflow rectangular gridded shapes represent matrices in which red squared shapes represent missing data points. To align with the spatial and temporal resolution of NDVI images, one must follow the pro-

cess until reaching the orange box. Similarly, for obtaining a 16-day hybrid dataset, the procedure should be pursued until reaching the blue box



Similarly, following the methodology outlined in Sect. 2.3.1, a 16-day hybrid dataset categorized by land cover types was generated. For each timestamp, this dataset encompassed NDVI, LST-Day ( $^{\circ}\text{C}$ ), LST-Night ( $^{\circ}\text{C}$ ), Relative Humidity Average (%), and Precipitation (mm) values for all classes, except Barren, Water Bodies, Permanent Snow and Ice, Urban and Built-up Lands, and Closed Shrublands. These classes were excluded due to either their low or nonexistent vegetation density, or the lack of significant distribution of such land cover types across the study area (Sulla-Menashe and Friedl 2018). This dataset was further employed for the subsequent analyses presented in the remainder of this article.

Normalized Cross-Correlation (NCC) is a widely applied mathematical concept found in diverse domains, such as signal processing, image analysis, and pattern recognition (Yoo and Han 2009; Kaso 2018; Pan et al. 2020). Its fundamental purpose is to assess the similarity between two sets of data or signals. In mathematical terms, the NCC between two signals, conventionally represented as  $x$  and  $y$ , is formally defined as follows (Yoo and Han 2009; Kaso 2018; Pan et al. 2020):

$$\text{NCC}(x, y) = \frac{\sum_{n=0}^{N-1} (x[n] - \bar{x})(y[n] - \bar{y})}{\sqrt{\sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \sum_{n=0}^{N-1} (y[n] - \bar{y})^2}}, \quad (1)$$

where,

- $N$  represents the length of the signals  $x$  and  $y$ .
- $x[n]$  and  $y[n]$  denote the discrete values of the two signals at time index  $n$ .
- $\bar{x}$  and  $\bar{y}$  are the means of the signals  $x$  and  $y$  respectively.

In this formulation, NCC quantifies the similarity between two signals by evaluating the normalized covariance of their deviations from their respective means. It yields a numerical value within the range of  $-1$  to  $1$ , where a result of  $1$  signifies a perfect positive correlation,  $-1$  indicates a perfect negative correlation, and  $0$  denotes no correlation between the signals (Yoo and Han 2009; Kaso 2018; Pan et al. 2020).

Determining time lags between two signals using NCC involved several essential steps. Initially, NCC values were computed for a range of potential lags by shifting one of the signals accordingly, creating an NCC time series. Subsequently, peaks in the NCC time series were identified, signifying significant correlations at specific lags. The position of these peaks denoted the direction and magnitude of the lag or lead between the signals. The lag with the highest NCC value was selected as the most prominent one (Yoo and Han 2009; Kaso 2018; Pan et al. 2020).

### 2.3.3 Trend Analyses

Seasonal and Trend decomposition using LOESS (STL) is a statistical methodology that was first introduced by Cleveland et al. (1990). Its primary objective is to break down time series data into three key components: trend, seasonality, and residuals. The trend component reflects the overall direction of the data over time, while the seasonal component captures cyclic patterns that repeat over time. The residual component represents the variation in the data that is not accounted for by the trend or seasonal components. Mathematically, a time series  $Y_t$  can be decomposed into its seasonality  $S_t$ , trend  $T_t$ , and residual  $R_t$  time series components using STL, as expressed in the following equation (Wen et al. 2019, 2020):

$$Y_t = T_t + S_t + R_t. \quad (2)$$

Wen et al. (2019) proposed RobustSTL as a more robust version of STL for time series decomposition. The RobustSTL algorithm can be used to decompose time series data by addressing the challenges of noise, outliers, and shifts in seasonality. The algorithm consists of four steps: denoising the time series, extracting the trend, calculating the seasonality component, and adjusting the extracted components. These steps are repeated until convergence, resulting in a robust decomposition of the time series data (Wen et al. 2019, 2020).

In time series analysis, it is essential to detect anomalous periods to comprehend unusual patterns or events that significantly deviate from the expected behavior of the data (Boschetti et al. 2013; Soriano-Vargas et al. 2021; Jamshidi et al. 2022). A powerful approach to identifying these periods is through the use of Z-score analysis on the trend component  $T_t$  obtained from RobustSTL. The Z-score analysis is a statistical technique that involves standardizing the data by subtracting the mean and dividing by the standard deviation. This standardization produces a score that indicates the number of standard deviations away from the mean that a particular data point is. In the context of time series analysis, the Z-score analysis of the trend component  $T_t$  can help pinpoint periods that significantly differ from the expected trend behavior. This approach can also assist in identifying potential outliers, providing critical insights into the underlying processes that drive the data (Boschetti et al. 2013; Soriano-Vargas et al. 2021; Jamshidi et al. 2022).

To initiate the process of detecting anomalous periods within a time series  $Y_t$ , the first step involves decomposing the series into its fundamental components—namely,  $S_t$ ,  $T_t$ , and  $R_t$ —utilizing the RobustSTL algorithm. Following this, the trend component  $T_t$  is independently evaluated, and instances that satisfy the following condition are identified (Zhang et al. 2007; Chakrabarti et al. 2008):

$$\begin{cases} \frac{T_t - \mu_t}{\sigma_t} \geq Z_{\alpha/2}, & \text{Indicating Significantly High,} \\ \frac{T_t - \mu_t}{\sigma_t} \leq -Z_{\alpha/2}, & \text{Indicating Significantly Low,} \end{cases} \quad (3)$$

where  $\mu_t$  and  $\sigma_t$  stand for the mean and standard deviation of  $T_t$ , respectively. Meanwhile,  $Z_{\alpha/2}$  represents the Z-score corresponding to the significance level  $\alpha$ . To illustrate, in cases where  $\alpha = 0.05$ ,  $Z_{\alpha/2}$  aligns with the 95th percentile of the standard normal distribution, providing a confidence level of 95% for the identification of these periods as anomalies.

Additionally, two non-parametric statistical methods, namely Mann–Kendall (MK) test (Mann 1945) and Sen's Slope Estimator (SSE) (Sen 1968), were utilized to analyze the data. The signs of the differences between observations were examined by the MK test to detect trends in the data. The magnitude and direction of the trends, indicating whether they indicated increasing or decreasing trends, were assessed by SSE (Ahmed and Hassan 2023).

The significance of the trends at the 95% and 99% confidence levels was determined by the MK test, which allowed the reliability of the findings to be assessed. These two methods enable trends in the data to be identified and evaluated without making assumptions about its underlying distribution, making them useful tools for the analysis (Ahmed and Hassan 2023).

To compute MK test statistic  $S$  for a time series  $x_1, x_2, \dots, x_n$  of length  $n$ , we use the following equation (Ahmed and Hassan 2023):

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_i - x_j), \quad (4)$$

where  $\text{sgn}(x_i - x_j)$  is the sign function defined as:

$$\text{sgn}(x_i - x_j) = \begin{cases} 1, & x_i - x_j > 0, \\ 0, & x_i - x_j = 0, \\ -1, & x_i - x_j < 0. \end{cases} \quad (5)$$

The sign function indicates whether the difference between the values of the time series at two distinct time points,  $i$  and  $j$ , is positive, negative, or zero.

The mean of variable  $S$  is zero, and the variance of  $S$  can be determined using the following equation:

$$\text{VAR}(S) = \frac{1}{18} \left( n(n-1)(2n+5) - \sum_{k=1}^p t_k(t_k-1)(2t_k+5) \right), \quad (6)$$

where  $p$  denotes the total number of tie groups in the data, and  $t_k$  represents the number of data points within the  $k$ -th tie group.

When the value of  $n$  is greater than 10, the standard normal test statistic  $Z$  is calculated based on the following conditions:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{VAR}(S)}}, & S > 0, \\ 0, & S = 0, \\ \frac{S+1}{\sqrt{\text{VAR}(S)}}, & S < 0. \end{cases} \quad (7)$$

where a positive or negative value of  $Z$  indicates an upward or downward trend, respectively.

The magnitude of trends was estimated using SSE, denoted by  $\beta$ , which is calculated using the following equation (Ahmed and Hassan 2023):

$$\beta = \text{Median} \left( \frac{x_j - x_i}{j - i} \right), \quad j > i, \quad (8)$$

where  $x_i$  and  $x_j$  represent the  $i$ -th and  $j$ -th ordered observations, and  $i$  and  $j$  are indices such that  $i < j$ . When the value of  $\beta$  is positive, it implies an upward trend during the period of interest, whereas a negative value of  $\beta$  indicates a downward trend (Ahmed and Hassan 2023). The 16-day hybrid dataset underwent a combined analysis of anomaly detection and the MK-trend test, facilitated by the application of the Sen's Slope Estimator (SSE).

### 2.3.4 Evaluating Climate's Role in Shaping Vegetation

CatBoost, a rapid and efficient machine learning algorithm (Prokhorenkova et al. 2018), was employed in our study. This algorithm operates by utilizing the technique of attribute splitting to delineate the feature space. It effectively mitigates gradient bias to mitigate the risk of overfitting. Furthermore, CatBoost employs a swift scoring mechanism, which leverages oblivious trees and feature binarization, thereby enhancing computational efficiency. This algorithm was employed to develop and train a set of 11 models using the hybrid 16-day dataset.

The process of identifying the most impactful climate variables for each LULC class was conducted systematically. Initially, a thorough examination encompassed several combinations of the following factors: NDVI, LST-Day ( $^{\circ}\text{C}$ ), LST-Night ( $^{\circ}\text{C}$ ), Relative Humidity Average (%), and Precipitation (mm). The selection of the most suitable features for each LULC class was determined by assessing the accuracy of initial models generated from all conceivable combinations of these variables (Buitinck et al. 2013).

The distribution of sizes between the training, validation, and test sets cannot be predetermined with complete accuracy and depends on the quantity of available data. This means that for smaller datasets, a greater proportion of the data will be used for validation and testing, while the training set will be relatively smaller. Conversely, as the size

of the dataset increases, the ratio of validation and testing data will decrease while the training data will increase, as explained in (Pawluszek-Filipiak and Borkowski 2020; Dastour and Hassan 2023). The training data for each model comprised all available information leading up to January 1st, 2019. For model evaluation, a distinct validation dataset, spanning from January 1st, 2019, to January 1st, 2021, was employed. Additionally, a separate testing dataset, encompassing information collected from January 1st, 2021, and onwards, was reserved exclusively to assess the models' performance.

The feature importance score generated by the CatBoost Regressor is a measure of how much each feature contributes to improving the model's accuracy when used in the decision trees. This score is typically normalized between 0 and 100, with 100 indicating that the feature is the most important. Feature importance refers to the degree to which each feature contributes to the model's ability to predict the target variables (Hancock and Khoshgoftaar 2020).

All trained models were subjected to evaluation utilizing metrics such as mean squared error (MSE), mean absolute error (MAE), and *R*-squared, encompassing assessments across the training, testing, and validation sets (Powers 2011; Dastour and Hassan 2023). Incorporating accuracy metrics into our evaluation process allows us to gauge the model's suitability for practical use. These metrics provide insights into the reliability and robustness of the feature importance results, which are pivotal for comprehending the connections between climate variables and LULC classes (Saarela and Jauhainen 2021).

### 3 Results

#### 3.1 Correlation Analyses

Figure 2 depicts pixel-wise linear correlation coefficients, computed at a 500 m spatial resolution, between NDVI and the following variables: LST-Day (°C), LST-Night (°C), Precipitation (mm), Relative Humidity Average (%), Digital Elevation Model (CDEM) of Alberta at a 30 m spatial resolution, and MODIS Land Cover Type-1 (LULC) data for 2021. This figure also displays linear correlations between NDVI and these variables across different land cover types, as determined using the 16-day hybrid dataset.

As depicted in Fig. 2a–f, the relationship between NDVI and both LST-Day and LST-Night appeared to have been influenced by elevation. Particularly, in regions covered by Grasslands, the correlation between NDVI and LST-Day seemed slightly diminished at higher altitudes compared to lower altitudes. Conversely, the correlation between NDVI and LST-Night was stronger at higher altitudes within Grassland areas compared to their lower-altitude counterparts.

Similar patterns were observed in areas characterized as Savannas. However, in Forested areas, the correlations between NDVI and LST-Day, as well as NDVI and LST-Night exhibited similar trends across both high and low altitudes.

In Fig. 2g, the correlation between NDVI and both LST-Day and LST-Night consistently exhibited high values across various land cover types. The lowest correlation was observed between NDVI and LST-Day in Cropland areas (0.88), while the highest was found in Savanna regions (0.93). Regarding the correlation between NDVI and LST-Night, the lowest values were observed in the Open Shrubland and Deciduous Needleleaf Forest areas (0.92 correlation), while higher values were found elsewhere (ranging from 0.93 to 0.95). Negative correlation values between NDVI and precipitation suggested a phase difference or time lag between them. When two-time series display a negative correlation, it generally indicates that as one series increases, the other tends to decrease, and vice versa. This behavior often suggests that one series is leading or lagging the other in time (Dean and Dunsmuir 2015; Zhou and Hua 2021).

Table 1 highlights the results of NCC analyses. In this table, the upper rows represent the time lag (in days) between NDVI and each variable. A negative lag indicates NDVI trailing behind, while a positive lag implies NDVI leading the variable. The second row displays the associated NCC values. Only NCC values exceeding the significance threshold of 0.7 have been incorporated into the table.

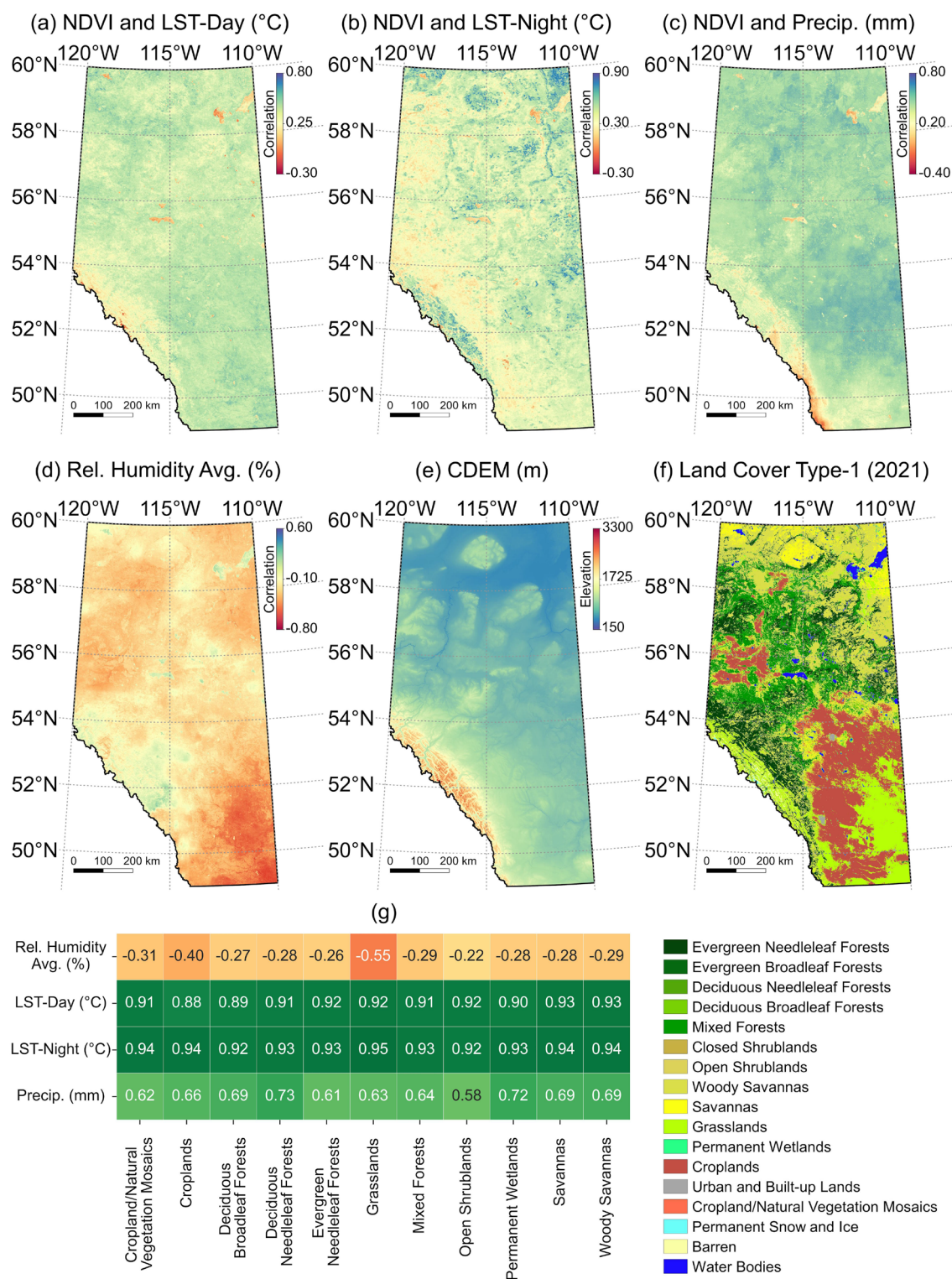
During the study period spanning 2001–2022, Table 1 illustrates the temporal dynamics. NDVI exhibited a lead over Relative Humidity Average of approximately 112 days (almost 3.5 months) across diverse land cover types. Additionally, LST-Day preceded NDVI by 16 days (approximately 2 weeks). Concerning LST-Night, a lead time of less than 16 days, indicated by zero values, was generally observed between NDVI and LST-Night for most land cover types. However, in regions characterized by Open Shrublands, Permanent Wetlands, Savannas, and Woody Savannas, NDVI was led by LST-Night by about two weeks. The relationship between NDVI and Precipitation revealed a consistent lag of NDVI behind Precipitation in most land cover types, with lead times ranging from 16 to 32 days (2 weeks to 1 month).

#### 3.2 Trend Analyses

Figure 3 provides a summary of the anomaly analysis conducted on the 16-day hybrid dataset. The figure shows the periods identified as anomalous based on the *Z*-score analysis of the trend component  $T_t$  obtained from RobustSTL.

The trend analysis of the 16-day hybrid dataset was performed using the MK test. To eliminate the influence of seasonality, the time series  $Y_t$  was decomposed into its





**Fig. 2** Pixel-wise linear correlations between **a** NDVI and LST-Day (°C), **b** NDVI and LST-Night (°C), **c** NDVI and precipitation (mm), **d** NDVI and relative humidity average (%), **e** The Canadian Digital Elevation Model (CDEM) and **f** Land Cover Type-1 (2021). **g** The

correlation of NDVI with relative humidity average (%), LST-Day (°C), LST-Night (°C), and precipitation (mm) categorized by land cover type



**Table 1** The table comprises pairs of rows, each conveying specific information

		CVM	C	DBF	DNF	ENF	G	MF	OS	PW	S	WS
Rel. humidity avg. (%)	Lag				– 128	– 112			– 112	– 128	– 112	– 112
	NCC				0.757	0.701			0.745	0.757	0.767	0.753
LST-Day (°C)	Lag	16	16	16	16	16	16	16	16	16	16	16
	NCC	0.940	0.898	0.919	0.931	0.924	0.931	0.929	0.936	0.937	0.952	0.95
LST-Night (°C)	Lag	0	0	0	16	0	0	0	16	16	16	16
	NCC	0.936	0.938	0.922	0.929	0.934	0.950	0.929	0.923	0.943	0.939	0.940
Precipitation (mm)	Lag		32	16	0		32			16		16
	NCC		0.726	0.703	0.727		0.700			0.751		0.703

The first row in each pair denotes the lag (in days) between NDVI and climate variables. A negative number for Lag means that the NDVI time series is leading the climate variable time series, and a positive Lag means the NDVI time series is lagging behind the climate variable time series. The second row presents the calculated NCC (normalized cross-correlation) values. Only NCC values surpassing the established threshold of 0.7 are incorporated in the table. Class abbreviations: Cropland/Natural Vegetation Mosaics (CVM), Croplands (C), Deciduous Broadleaf Forests (DBF), Deciduous Needleleaf Forests (DNF), Evergreen Needleleaf Forests (ENF), Grasslands (G), Mixed Forests (MF), Open Shrublands (OS), Permanent Wetlands (PW), Savannas (S), and Woody Savannas (WS)

trend ( $T_t$ ), seasonal ( $S_t$ ), and residual ( $R_t$ ) components using the RobustSTL method. Subsequently, the trend component  $T_t$  was subjected to the MK trend analysis. The results of the trend analysis on this dataset are presented in Table 2.

The results of the Mann–Kendall test and Sen’s Slope analysis conducted on the annual percentage of LULC classes from 2001 to 2022 are depicted in Fig. 4. This analysis yielded insights into the temporal trends in the percentage of LULC classes and facilitated the identification of significant changes in land use patterns.

Figure 4b illustrated an increasing trend in Croplands area percentage over the study period, with a slope of 0.066. For Cropland areas, Table 2 demonstrated a similar increasing trend in relative humidity with a slope of 0.0053. In Fig. 4b, lower Cropland coverage was exhibited during the years 2001 and 2002 when compared to the estimated trendline. This observation was in alignment with the findings presented in Fig. 3b, wherein significantly lower values of Relative Humidity Average, LST-Day, LST-Night, and Precipitation were observed during the years 2001 and 2002, in contrast to the remaining duration of the study period.

Figure 4d depicted an upward trend in the percentage of Evergreen Needleleaf forest area over the study period, characterized by a slope of 0.057. A corresponding increase in precipitation (slope 0.01) for areas covered by Evergreen Needleleaf forests was revealed in Table 2. Specifically, during the years 2001 and 2002, the coverage of the Evergreen Needleleaf forest surpassed that of the estimated trendline. This observation aligns with the observations presented in Fig. 3e, wherein significantly higher values for NDVI were evident. Similarly, the year 2006 is notable in Fig. 4d due to slightly lower coverage when compared to the estimated trendline. This observation is in concordance with the data presented in Fig. 3e, which exhibited significantly higher

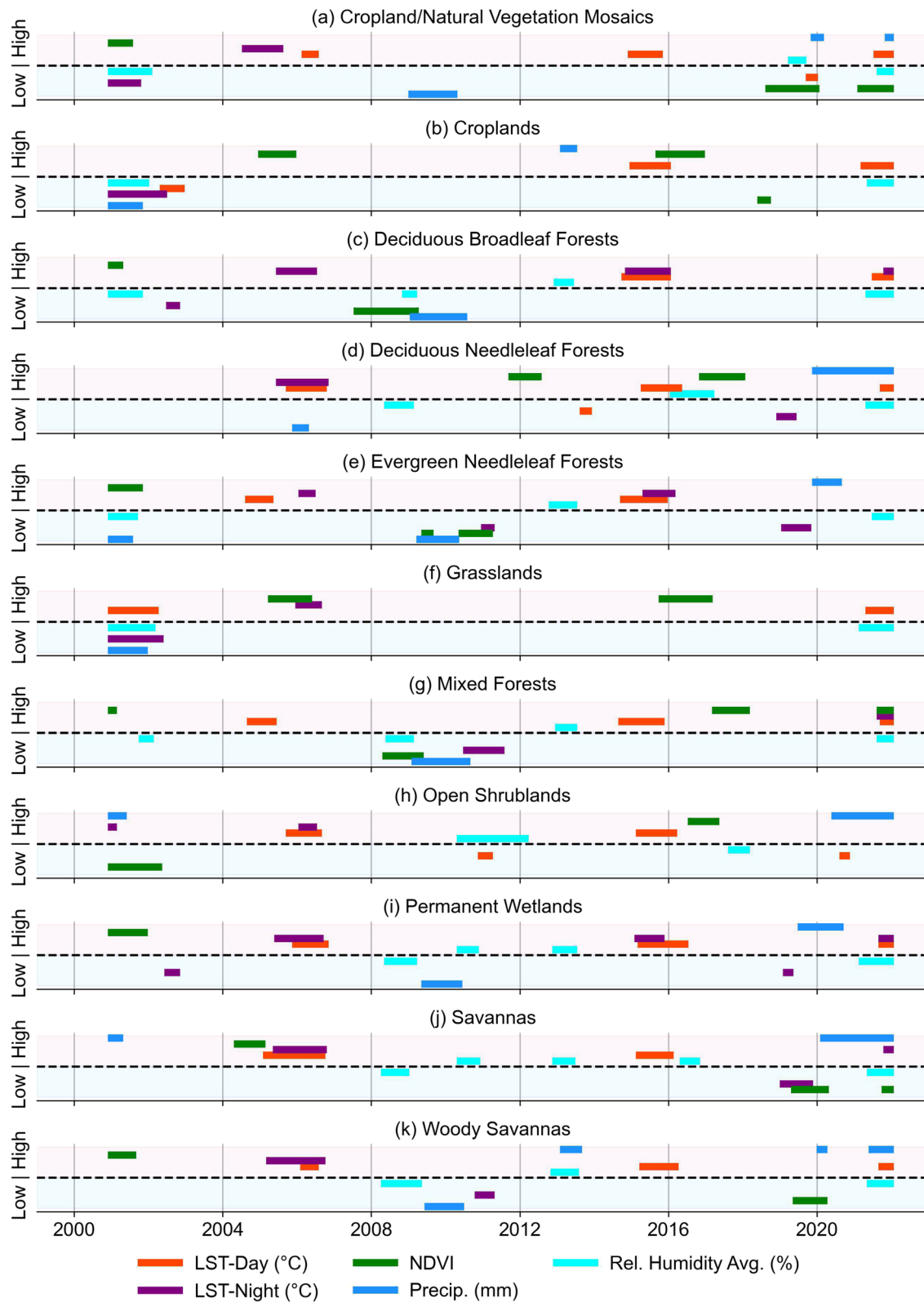
values of LST-Day and LST-Night around 2006, as contrasted with the remainder of the study period.

In the case of Grasslands, a declining trend in overall coverage percentage was unveiled by Fig. 4e, marked by a slope of  $-0.112$ . Table 2 indicated that during the 2001–2022 period, an upward slope of 0.0019 was observed for Relative Humidity Average and 0.0035 for Precipitation, while LST-Day exhibited a declining trend with a slope of  $-0.0019$ . Furthermore, Fig. 4e also indicated that the coverage percentages for 2000 and 2001 were higher compared to the estimated trendline. Additionally, Fig. 3f highlighted values of Relative Humidity Average, LST-Night, and Precipitation during 2001 and 2002 that were significantly lower, accompanied by LST-Day values that were significantly higher, contrasting with the rest of the study period.

Figure 4i evidenced an increase in coverage percentage for Woody Savannas around 2006, surpassing the estimated trendline. Simultaneously, Fig. 3k unveiled significantly higher values for LST-Day and LST-Night around the same year. In terms of trend analysis, Fig. 4i delineated a decrease in the percentage of areas covered by Woody Savannas, characterized by a slope of  $-0.141$ . Complementarily, Table 2 indicated declining trends in LST-Night and NDVI.

### 3.3 Evaluating Climate’s Role in Shaping Vegetation

The accuracy results of the CatBoost models for NDVI, which were based on Relative Humidity Average, LST-Day, LST-Night, and Precipitation, are presented in Table 3. The models demonstrated a great level of accuracy, particularly when assessed on the validation set, which represents unseen data from the training and validation sets. The  $R$ -squared accuracy scores ranged from 0.931 to 0.983 on the test set, showing the efficacy of the models. This high accuracy enhances both the selection of influential climate variables



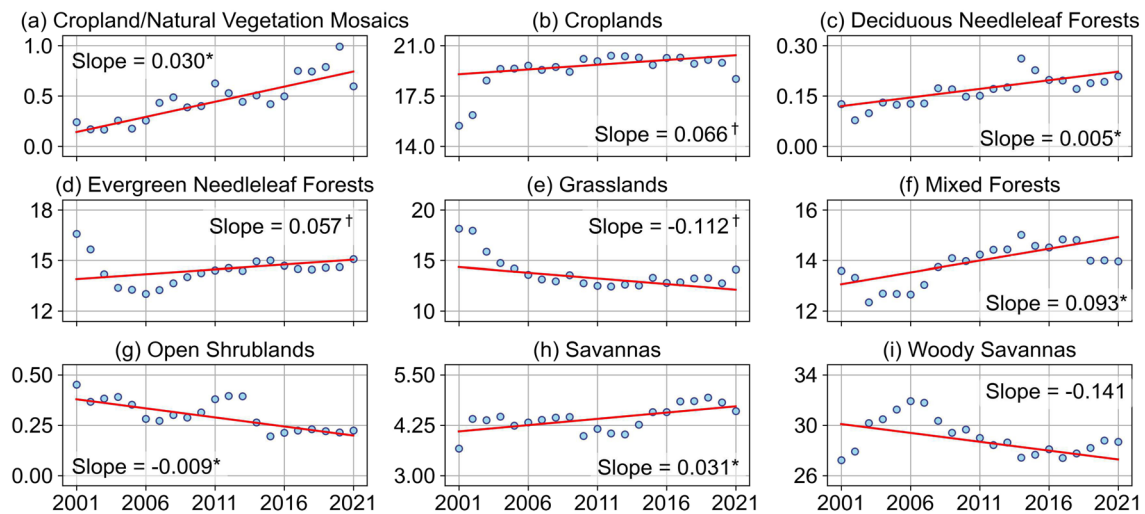
**Fig. 3** Anomaly periods were identified in each time series using Z-score analysis of the trend component  $T_i$  obtained through Robust-STL, with a significance level of  $\alpha = 0.05$ . The x-axis spans from 2000 to 2022 in all panels, with horizontal lines denoting anomalous

periods. Using Eq. (3), periods with significantly high values are shaded pink at the top of each panel, while periods with significantly low values are shaded blue at the bottom

**Table 2** Mann–Kendall test and Sen's slope were used to analyze trend component ( $T_t$ ) of various time series, obtained through RobustSTL, with significant slopes shown at 90% confidence level or higher (\* for 99%, † for 95%)

	CVM	C	DBF	DNF	ENF	G
Rel. humidity avg. (%)	+ 4.71E−03*	+ 5.32E−03*	+ 4.15E−03*	9.14E−04	+ 6.35E−03*	+ 1.89E−03*
LST-Day (°C)	− 8.31E−04*			− 1.10E−03*	− 4.00E−04	− 1.94E−03*
LST-Night (°C)			+ 3.42E−04†	− 1.27E−03*	− 5.00E−04*	
NDVI	− 9.75E−05*			+ 1.56E−05*	− 5.22E−05*	− 1.34E−05*
Precipitation (mm)	+ 8.22E−03*		+ 4.98E−03*	+ 9.41E−03*	+ 1.01E−02*	+ 3.51E−03*
	MF	OS	PW	S	WS	
Rel. humidity avg. (%)	+ 7.54E−03*	− 1.39E−03*	+ 1.07E−03*	+ 1.30E−03†	+ 3.26E−03*	
LST-Day (°C)		− 7.09E−04*		− 1.58E−03*		
LST-Night (°C)			− 3.72E−04	− 9.07E−04*	− 3.73E−04	
NDVI	+ 3.97E−05*	+ 4.91E−05*	− 4.63E−05*	− 3.56E−05*	− 2.68E−05*	
Precipitation (mm)	+ 6.68E−03*	+ 5.67E−03*	+ 3.44E−03*	+ 3.18E−03*	+ 4.37E−03*	

Class abbreviations: Cropland/Natural Vegetation Mosaics (CVM), Croplands (C), Deciduous Broadleaf Forests (DBF), Deciduous Needleleaf Forests (DNF), Evergreen Needleleaf Forests (ENF), Grasslands (G), Mixed Forests (MF), Open Shrublands (OS), Permanent Wetlands (PW), Savannas (S), and Woody Savannas (WS)

**Fig. 4** Mann–Kendall test and Sen's Slope were used to analyze the percentage of LULC classes over the span of 2001–2022, with significant slopes shown at 90% confidence level or higher (\* for 99%, † for 95%)

and confidence in the reliability of feature importance results.

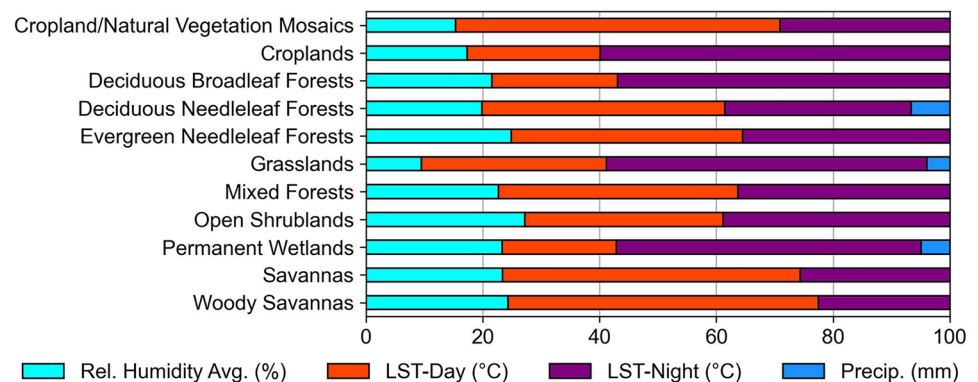
Figure 5 presents a summary of feature importance for 11 models that were developed using the CatBoost Regressor to characterize NDVI in relation to four variables: Relative Humidity Average (%), LST-Day (°C), LST-Night (°C), and Precipitation (mm). Notably, a prominent role was played by the LST-Day and LST-Night variables in the NDVI modeling process when utilizing the 16-day hybrid dataset. Specifically, an average importance of 37.42% was

exhibited for LST-Day, while LST-Night demonstrated an average importance of 40.35%. These findings suggested a substantial influence of temperature-related attributes in the modeling of NDVI, aligning with results that were obtained from Correlation Analyses and Trend Analyses. Conversely, Precipitation was found to have the least impact and was exclusively employed in NDVI modeling across Deciduous Needleleaf Forests, Grasslands, and Permanent Wetlands. Its average importance in these specific contexts was 5.15%.

**Table 3** Accuracy results of CatBoost regressor models on train, validation, and test sets for NDVI by each LULC class

LULC class	Train			Validation			Test		
	MSE	MAE	R <sup>2</sup>	MSE	MAE	R <sup>2</sup>	MSE	MAE	R <sup>2</sup>
Croplands	0.002	0.032	0.969	0.004	0.051	0.944	0.003	0.045	0.932
Cropland/Natural Vegetation Mosaics	0.003	0.039	0.966	0.007	0.055	0.934	0.005	0.042	0.941
Deciduous Broadleaf Forests	0.002	0.035	0.971	0.004	0.043	0.955	0.002	0.037	0.973
Deciduous Needleleaf Forests	0.002	0.028	0.980	0.004	0.045	0.953	0.002	0.037	0.970
Evergreen Needleleaf Forests	0.002	0.029	0.955	0.002	0.038	0.942	0.002	0.029	0.957
Grasslands	0.001	0.029	0.961	0.002	0.033	0.955	0.002	0.036	0.931
Mixed Forests	0.003	0.039	0.960	0.003	0.035	0.957	0.002	0.035	0.965
Open Shrublands	0.001	0.023	0.985	0.003	0.033	0.969	0.003	0.032	0.969
Permanent Wetlands	0.001	0.020	0.984	0.002	0.029	0.963	0.001	0.023	0.983
Savannas	0.001	0.025	0.982	0.003	0.033	0.966	0.002	0.033	0.972
Woody Savannas	0.002	0.026	0.977	0.002	0.034	0.970	0.003	0.039	0.965

The mean squared error (MSE), mean absolute error (MAE), and  $R^2$  ( $R$ -squared) respectively measure the discrepancy between predicted and actual values. Smaller values of MSE and MAE are indicative of enhanced accuracy. The maximum achievable accuracy for  $R$ -squared is 1

**Fig. 5** Feature Importance Analysis for NDVI Modeling which summarizes the feature importance results obtained from 11 models developed using the CatBoost Regressor for modeling NDVI. The analysis focuses on the influence of four environmental variables: Relative Humidity Average (%), LST-Day (°C), LST-Night (°C), and precipitation (mm)

## 4 Discussion

The relationship between NDVI and both LST-Day and LST-Night was found to be influenced by elevation. Specifically, in regions covered by Grasslands, it was observed that the correlation between NDVI and LST-Day was somewhat diminished at higher altitudes when compared to lower altitudes. Conversely, the correlation between NDVI and LST-Night exhibited greater strength at higher altitudes within Grassland areas in contrast to their lower-altitude counterparts. Similar trends were discerned in areas characterized as Savannas. However, in forested areas, the correlations between NDVI and both LST-Day and LST-Night demonstrated consistent patterns across both high and low altitudes. This elevation-mediated effect underscores the importance of considering local topographical features when assessing the impact of climate on vegetation, as higher altitudes may exhibit distinct responses to temperature variations. Several studies have argued that higher altitudes may exhibit

different responses to temperature fluctuations compared to lower altitudes, contingent upon the prevailing land cover types (Karnieli et al. 2019; Huang et al. 2020; Lu et al. 2023; Dewan and Corner 2012).

Normalized Cross-Correlation (NCC) analyses were conducted to explore the temporal dynamics of NDVI and climate variables. The analysis revealed that NDVI consistently preceded the Relative Humidity Average by about 3.5–4 months across Deciduous Needleleaf Forests (DNF), Evergreen Needleleaf Forests (ENF), Open Shrublands (OS), Permanent Wetlands (PW), Savannas (S), and Woody Savannas (WS). This implies that NDVI could affect the relative humidity with a seasonal lag, as NDVI could alter the soil moisture and the water stress of plants. Higher NDVI values could signal increased vegetation growth in the next season, while lower NDVI values could indicate decreased vegetation growth. This suggests a seasonal gap between vegetation and humidity. Other studies have also reported similar findings (Han et al. 2023; Chanda et al. 2024; Xie et al. 2023; Brindle et al. 2023). On the other hand, LST-Day exhibited



a lead over NDVI of approximately two weeks, indicating a short-term temporal influence of daytime temperature patterns on NDVI. For most land cover types, LST-Night demonstrated minimal temporal lag with NDVI. However, in regions characterized by lower-density vegetation, such as Open Shrublands, Permanent Wetlands, Savannas, and Woody Savannas, LST-Night led NDVI by approximately two weeks. The short-term lag in the influence of LST-Day on NDVI could be attributed to its reflection of solar radiation and the photosynthetic activity of plants. Higher LST-Day appears to enhance NDVI in the subsequent weeks, while lower LST-Day has the opposite effect. These findings align with similar observations reported in other studies (Dastour et al. 2022; Wang et al. 2003). NDVI exhibited a temporal lag of 2–4 weeks behind Precipitation across diverse land cover types, indicating a pattern of vegetation greenness change following precipitation events within the same timeframe. Precipitation leads NDVI by 2–4 weeks because precipitation provides water and nutrients for plant growth, and reduces soil moisture stress and drought (Soomro et al. 2021). These findings align with previous studies (Dastour et al. 2022; Wang et al. 2003; Ji and Peters 2005; Ndehedehe et al. 2021; Yuan et al. 2015; Garai et al. 2022). The results showed that different climate variables influenced NDVI with different temporal lags, depending on the land cover type. This means that the vegetation response to climate change was not uniform across the region, but varied according to the characteristics of the vegetation and the local climate conditions.

Significant trends in land cover dynamics were discerned throughout the study period. Notably, Cropland areas, covering an average of just under 20% of Alberta during 2001–2022, manifested a consistent increase in coverage, paralleling an ascending trend in relative humidity. However, during the years 2001 and 2002, Cropland coverage deviated below the projected trendline, coinciding with a substantial reduction in humidity, LST-Day, LST-Night, and precipitation. In contrast, Evergreen Needleleaf forest areas, which encompassed approximately 14.5% of Alberta during 2001–2022, exhibited an upward trend in coverage, concomitant with increased precipitation. Notably, the years 2001 and 2002 featured higher Evergreen Needleleaf forest coverage, corresponding to significantly high NDVI values. Conversely, a minor dip in coverage around 2006 aligned with significantly high LST-Day and LST-Night values. Likewise, Grasslands, comprising about 13.8% of Alberta during 2001–2022, experienced a diminishing coverage trend, notwithstanding increasing humidity and precipitation, yet accompanied by declining LST-Day. Notable higher coverage percentages in 2000 and 2001 coincided with markedly lower humidity, LST-Night, and precipitation. Lastly, Woody Savannas, accounting for roughly 29% of Alberta, on average, during 2001 and 2022, displayed

increased coverage around 2006, surpassing the trendline, while concurrently exhibiting elevated LST-Day and LST-Night values. Nevertheless, over the study period, a declining trend in Woody Savanna's coverage was evident, aligning with decreasing LST-Night and NDVI values. These trends provide insights into the complex interplay between land cover changes and climatic factors in Alberta, emphasizing the sensitivity of different land cover types to variations in humidity and temperature.

NDVI may have been influenced by LST-Day and LST-Night, with average importance values of 37.42% and 40.35%, respectively. This shows the significant role of temperature variables in NDVI modeling, consistent with the results from Correlation Analyses and Trend Analyses. On the other hand, Precipitation may have had a much lower impact and was only used in NDVI modeling for Deciduous Needleleaf Forests, Grasslands, and Permanent Wetlands, with an average importance value of 5.15%. The variations in NDVI across diverse land cover types could potentially be predominantly influenced by temperature-related factors, thereby playing a crucial role in comprehending vegetation dynamics. The significance of LST-Day and LST-Night as a key climate variables in NDVI modeling suggested the relevance of temperature-related factors in shaping vegetation patterns across various land cover types. This implication could underscore the ecological importance of temperature regulation within Alberta's ecosystems.

Atmospheric conditions (e.g., clouds, dust, fog, smoke) can degrade the quality and accuracy of the remote sensing data, which is one of the limitations of this study. Another limitation is the low resolution (spatial, temporal, or spectral) of the remote sensing data, which can limit its representation of the object or the environment in detail, change, or variation. Moreover, the study only spanned a 22-year period, and more data and a longer-term analysis could potentially improve the results.

## 5 Conclusions

In summary, this study has provided clarity on the intricate relationship between the Normalized Difference Vegetation Index (NDVI) and key environmental variables, namely Land Surface Temperature (LST) during the daytime (LST-Day) and nighttime (LST-Night), along with Precipitation, across diverse land cover types in the Alberta region.

Elevation affects the correlation between NDVI and both LST-Day and LST-Night depending on the land cover type. In Grasslands and Savannas, higher altitudes weaken the NDVI-LST-Day correlation but strengthen the NDVI-LST-Night correlation. In forests, the NDVI-LST correlations are consistent across altitudes. This shows that local topography influences how vegetation responds to climate, as higher

altitudes may have different temperature sensitivities than lower altitudes.

Temporal dynamics were investigated using Normalized Cross-Correlation (NCC) analyses. The examination indicated a lead time of approximately 3.5–4 months for the NDVI preceding the Relative Humidity Average. This pattern was observed across various ecosystems, namely Deciduous Needleleaf Forests (DNF), Evergreen Needleleaf Forests (ENF), Open Shrublands (OS), Permanent Wetlands (PW), Savannas (S), and Woody Savannas (WS). LST-Day exhibited a lead over NDVI of about 2 weeks, indicating a short-term temporal influence of daytime temperature patterns on NDVI. NDVI exhibited a temporal lag of 2–4 weeks behind Precipitation across diverse land cover types.

Significant trends in land cover dynamics were observed in Alberta from 2001 to 2022. Cropland areas, constituting nearly 20% of the region, consistently increased alongside rising relative humidity. However, deviations occurred in 2001 and 2002, aligning with reductions in humidity, LST-Day, LST-Night, and precipitation. Evergreen Needleleaf forests, covering approximately 14.5%, exhibited an upward trend linked to increased precipitation, with notable peaks in 2001 and 2002 corresponding to high NDVI values. Grasslands, comprising 13.8%, showed diminishing coverage despite rising humidity and precipitation but accompanied by declining LST-Day. Woody Savannas, accounting for around 29%, displayed increased coverage in 2006, surpassing the trendline, with elevated LST-Day and LST-Night values. Despite this, a declining trend in Woody Savanna's coverage was evident, correlating with decreasing LST-Night and NDVI values over the study period. These findings underscore the intricate relationship between land cover changes and climatic factors in Alberta, emphasizing the sensitivity of various land cover types to variations in humidity and temperature.

Temperature variables (LST-Day and LST-Night) had high importance values (37.42% and 40.35%) in NDVI modeling, matching the Correlation and Trend Analyses results. Precipitation had a low importance value (5.15%) and was only used for some land cover types. NDVI variations across land cover types could be mainly driven by temperature-related factors, which are important for understanding vegetation dynamics and patterns. This implies that temperature regulation is ecologically vital for Alberta's ecosystems.

The study investigated how different vegetation types, elevation, and climate variables interacted to affect the Normalized Difference Vegetation Index (NDVI), a measure of vegetation health and productivity. The study could provide valuable information for conservation, impact assessment, and decision support for climate adaptation and mitigation.

Future research directions could include exploring the potential impacts of climate change on vegetation dynamics in Alberta, especially in relation to the frequency and

intensity of droughts and wildfires. Additionally, more studies are needed to investigate the interactions between vegetation and other biophysical factors, such as soil moisture, albedo, and carbon sequestration. Furthermore, the application of machine learning techniques, such as CatBoost, could be extended to other regions and ecosystems, as well as to multi-temporal and multi-spectral data sources, to improve the accuracy and robustness of NDVI modeling and its interpretations.

**Acknowledgements** The following entities are gratefully acknowledged by the authors: (i) the Government of Alberta (Alberta Agriculture and Forestry) for providing interpolated weather data for the Alberta townships for non-commercial/research purposes, and (ii) NASA for providing the MODIS NDVI imagery, MODIS LST-Day, MODIS LST-Night, and MODIS Land Cover Type-1 datasets. The authors also acknowledge the efforts of the Google personnel who made these data sets available on the Google Earth Engine.

**Funding** No funding was received.

**Data availability** The data used in this research are available in the public domain.

## Declarations

**Conflict of Interest** The authors declare no conflicts of interest.

## References

- Agriculture and Irrigation Department (2023) Current and historical Alberta weather station data viewer. <https://acis.alberta.ca/weather-data-viewer.jsp>. [Online]. Accessed 01 Feb 2023
- Ahmed MR, Hassan QK (2023) Occurrence, area burned, and seasonality trends of forest fires in the natural subregions of Alberta over 1959–2021. *Fire* 6(3):96. <https://doi.org/10.3390/fire6030096>
- Al Balasmeh OI, Karmaker T (2019) Effect of temperature and precipitation on the vegetation dynamics of high and moderate altitude natural forests in India. *J Indian Soc Remote Sens* 48(1):121–144. <https://doi.org/10.1007/s12524-019-01065-8>
- Alba A, Viguera-Gomez JF, Arce-Santana ER et al (2015) Phase correlation with sub-pixel accuracy: a comparative study in 1D and 2D. *Comput Vis Image Underst* 137:76–87. <https://doi.org/10.1016/j.cviu.2015.03.011>
- Ali M (2013) Climate change impacts on plant biomass growth. Springer, Dordrecht. <https://doi.org/10.1007/978-94-007-5370-9>
- Baldwin K, Allen L, Basquill S et al (2020) Vegetation zones of Canada: a biogeoclimatic perspective; information report. Technical report, GLC-X-25
- Barber QE, Nielsen SE, Hamann A (2015) Assessing the vulnerability of rare plants using climate change velocity, habitat connectivity, and dispersal ability: a case study in Alberta, Canada. *Reg Environ Change* 16(5):1433–1441. <https://doi.org/10.1007/s10113-015-0870-6>
- Bari E, Nipa NJ, Roy B (2021) Association of vegetation indices with atmospheric & biological factors using MODIS time series products. *Environ Chall* 5:100376. <https://doi.org/10.1016/j.envc.2021.100376>
- Bellard C, Bertelsmeier C, Leadley P et al (2012) Impacts of climate change on the future of biodiversity. *Ecol Lett* 15(4):365–377. <https://doi.org/10.1111/j.1461-0248.2011.01736.x>

- Boschetti M, Nutini F, Brivio PA et al (2013) Identification of environmental anomaly hot spots in West Africa from time series of NDVI and rainfall. *ISPRS J Photogramm Remote Sens* 78:26–40. <https://doi.org/10.1016/j.isprsjprs.2013.01.003>
- Brindle HE, Bastos LS, Christley R et al (2023) The spatio-temporal distribution of acute encephalitis syndrome and its association with climate and landcover in Vietnam. *BMC Infect Dis* 23(1):1–15. <https://doi.org/10.1186/s12879-023-08300-1>
- Buitinck L, Louppe G, Blondel M et al (2013) API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD workshop: languages for data mining and machine learning*, pp 108–122. [https://dai.cs.kuleuven.be/events/lml2013/papers/lml2013\\_api\\_sklearn.pdf](https://dai.cs.kuleuven.be/events/lml2013/papers/lml2013_api_sklearn.pdf)
- Chakrabarti S, Neapolitan R, Pyle D et al (2008) Data mining: know it all. Elsevier Science, Amsterdam. <https://books.google.ca/books?id=WRqZ0QsdXKkC>
- Chanda R, Singh SS, Singh NS et al (2024) Two-decadal climate impacts on growth of major forest types of Eastern Himalaya. *Trees For People*. <https://doi.org/10.2139/ssrn.4639163>
- Chen J, Yan F, Lu Q (2020) Spatiotemporal variation of vegetation on the Qinghai–Tibet Plateau and the influence of climatic factors and human activities on vegetation trend (2000–2019). *Remote Sens* 12(19):3150. <https://doi.org/10.3390/rs12193150>
- Chen A, Lantz TC, Hermosilla T, Wulder MA (2021) Biophysical controls of increased tundra productivity in the western Canadian Arctic. *Remote Sens Environ* 258:112358. <https://doi.org/10.1016/j.rse.2021.112358>
- Cleveland RB, Cleveland WS, McRae JE et al (1990) STL: a seasonal-trend decomposition. *J Off Stat* 6(1):3–73
- Cui L, Wang L, Singh RP et al (2018) Association analysis between spatiotemporal variation of vegetation greenness and precipitation/temperature in the Yangtze River Basin (China). *Environ Sci Pollut Res* 25(22):21867–21878. <https://doi.org/10.1007/s11356-018-2340-4>
- Dastour H, Hassan QK (2023) A comparison of deep transfer learning methods for land use and land cover classification. *Sustainability* 15(10):7854. <https://doi.org/10.3390/su15107854>
- Dastour H, Ghaderpour E, Zaghloul MS et al (2022) Wavelet-based spatiotemporal analyses of climate and vegetation for the Athabasca river basin in Canada. *Int J Appl Earth Observ Geoinf* 114:103044. <https://doi.org/10.1016/j.jag.2022.103044>
- Dean RT, Dunsmuir WTM (2015) Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: the importance of constructing transfer function autoregressive models. *Behav Res Methods* 48(2):783–802. <https://doi.org/10.3758/s13428-015-0611-2>
- Dewan AM, Corner RJ (2012) The impact of land use and land cover changes on land surface temperature in a rapidly urbanizing megacity. In: *2012 IEEE international geoscience and remote sensing symposium*. IEEE, pp 6337–6339. <https://doi.org/10.1109/IGARSS.2012.6352709>
- Downing D, Pettapiece W (2006) Natural regions and subregions of Alberta, Natural Regions Committee. Government of Alberta, Alberta
- Estevo CA, Stralberg D, Nielsen SE et al (2022) Topographic and vegetation drivers of thermal heterogeneity along the boreal-grassland transition zone in western Canada: implications for climate change refugia. *Ecol Evol*. <https://doi.org/10.1002/ece3.9008>
- Gao Y, Zhou X, Wang Q et al (2013) Vegetation net primary productivity and its response to climate change during 2001–2008 in the Tibetan Plateau. *Sci Total Environ* 444:356–362. <https://doi.org/10.1016/j.scitotenv.2012.12.014>
- Garai S, Khatun M, Singh R et al (2022) Assessing correlation between rainfall, normalized difference vegetation index (NDVI) and land surface temperature (LST) in Eastern India. *Saf Extrem Environ* 4(2):119–127. <https://doi.org/10.55779/ng2353>
- Grimm NB, Chapin FS III, Bierwagen B et al (2013) The impacts of climate change on ecosystem structure and function. *Front Ecol Environ* 11(9):474–482. <https://doi.org/10.1890/120282>
- Guo P, Zhao X, Shi J et al (2020) The influence of temperature and precipitation on the vegetation dynamics of the tropical island of Hainan. *Theor Appl Climatol* 143(1–2):429–445. <https://doi.org/10.1007/s00704-020-03430-x>
- Halofsky JS, Conklin DR, Donato DC et al (2018) Climate change, wildfire, and vegetation shifts in a high-inertia forest landscape: Western Washington, USA. *PLoS One* 13(12):e0209490. <https://doi.org/10.1371/journal.pone.0209490>
- Han Y, Lin Y, Zhou P et al (2023) Dynamic change, driving mechanism and spatiotemporal prediction of the normalized vegetation index: a case study from Yunnan province, China. *Front Ecol Evol*. <https://doi.org/10.3389/fevo.2023.1177849>
- Hancock JT, Khoshgofaar TM (2020) CatBoost for big data: an interdisciplinary review. *J Big Data*. <https://doi.org/10.1186/s40537-020-00369-8>
- Hassan Q, Bourque C (2010) Spatial enhancement of MODIS-based images of leaf area index: application to the boreal forest region of Northern Alberta, Canada. *Remote Sens* 2(1):278–289. <https://doi.org/10.3390/rs2010278>
- Huang S, Tang L, Hupy JP et al (2020) A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J For Res* 32(1):1–6. <https://doi.org/10.1007/s11676-020-01155-1>
- Hussien K, Kebede A, Mekuriaw A et al (2023) Spatiotemporal trends of NDVI and its response to climate variability in the Abbay River Basin, Ethiopia. *Heliyon* 9(3):e14113. <https://doi.org/10.1016/j.heliyon.2023.e14113>
- Jamshidi EJ, Yusup Y, Kayode JS et al (2022) Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: a case study on surface water temperature. *Ecol Inform* 69:101672. <https://doi.org/10.1016/j.ecoinf.2022.101672>
- Ji L, Peters AJ (2005) Lag and seasonality considerations in evaluating AVHRR NDVI response to precipitation. *Photogramm Eng Remote Sens* 71(9):1053–1061. <https://doi.org/10.1016/j.heliyon.2023.e20518>
- Jia L, Li Z, Xu G et al (2020) Dynamic change of vegetation and its response to climate and topographic factors in the Xijiang River basin, China. *Environ Sci Pollut Res* 27(11):11637–11648. <https://doi.org/10.1007/s11356-020-07692-w>
- Jiang R, Xie J, He H et al (2016) Spatiotemporal variability and predictability of normalized difference vegetation index (NDVI) in Alberta, Canada. *Int J Biometeorol* 60(9):1389–1403. <https://doi.org/10.1007/s00484-015-1132-5>
- Julien Y, Sobrino JA (2009) The yearly land cover dynamics (YLCD) method: an analysis of global vegetation from NDVI and LST parameters. *Remote Sens Environ* 113(2):329–334. <https://doi.org/10.1016/j.rse.2008.09.016>
- Karnieli A, Ohana-Levi N, Silver M et al (2019) Spatial and seasonal patterns in vegetation growth-limiting factors over Europe. *Remote Sens* 11(20):2406. <https://doi.org/10.3390/rs11202406>
- Kaso A (2018) Computation of the normalized cross-correlation by fast Fourier transform. *PLoS One* 13(9):e0203434. <https://doi.org/10.1371/journal.pone.0203434>
- Lee Rodgers J, Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. *Am Stat* 42(1):59–66. <https://doi.org/10.1080/00031305.1988.10475524>
- Liu Z, Menzel L (2016) Identifying long-term variations in vegetation and climatic variables and their scale-dependent relationships: a case study in Southwest Germany. *Glob Planet Change* 147:54–66. <https://doi.org/10.1016/j.gloplacha.2016.10.019>



- Lu L, Fu P, Dewan A et al (2023) Contrasting determinants of land surface temperature in three megacities: implications to cool tropical metropolitan regions. *Sustain Cities Soc* 92:104505. <https://doi.org/10.1016/j.scs.2023.104505>
- Mann HB (1945) Nonparametric tests against trend. *Econom: J Econom Soc*. <https://doi.org/10.2307/1907187>
- Ndehedehe CE, Ferreira VG, Agutu NO et al (2021) What if the rains do not come? *J Hydrol* 595:126040. <https://doi.org/10.1016/j.jhydrol.2021.126040>
- Pan C, Lv Z, Hua X et al (2020) The algorithm and structure for digital normalized cross-correlation by using first-order moment. *Sensors* 20(5):1353. <https://doi.org/10.3390/s20051353>
- Pawluszek-Filipiak K, Borkowski A (2020) On the importance of train-test split ratio of datasets in automatic landslide detection by supervised classification. *Remote Sens* 12(18):3054. <https://doi.org/10.3390/rs12183054>
- Powers D (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol* 2(1):37–63
- Prevedello JA, Winck GR, Weber MM et al (2019) Impacts of forestation and deforestation on local temperature across the globe. *PLoS One* 14(3):e0213368. <https://doi.org/10.1371/journal.pone.0213368>
- Prokhorenkova L, Gusev G, Vorobev A et al (2018) CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems* 31 (NeurIPS 2018). [https://papers.nips.cc/paper\\_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html](https://papers.nips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html)
- Saarela M, Jauhiainen S (2021) Comparison of feature importance measures as explanations for classification models. *SN Appl Sci*. <https://doi.org/10.1007/s42452-021-04148-9>
- Sekhon NS, Hassan QK, Sleep RW (2010) Evaluating potential of MODIS-based indices in determining “snow gone” stage over forest-dominant regions. *Remote Sens* 2(5):1348–1363. <https://doi.org/10.3390/rs2051348>
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc* 63(324):1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Soomro S, Hu C, Jian S et al (2021) Precipitation changes and their relationships with vegetation responses during 1982–2015 in Kunhar River basin, Pakistan. *Water Supply* 21(7):3657–3671. <https://doi.org/10.2166/ws.2021.129>
- Soriano-Vargas A, Werneck R, Moura R et al (2021) A visual analytics approach to anomaly detection in hydrocarbon reservoir time series data. *J Pet Sci Eng* 206:108988. <https://doi.org/10.1016/j.petrol.2021.108988>
- Stralberg D, Wang X, Parisien MA et al (2018) Wildfire-mediated vegetation change in boreal forests of Alberta, Canada. *Ecosphere* 9(3):e02156. <https://doi.org/10.1002/ecs2.2156>
- Sulla-Menashé D, Friedl MA (2018) User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product, vol 1. USGS, Reston, p 18
- Sun Q, Liu W, Gao Y et al (2020) Spatiotemporal variation and climate influence factors of vegetation ecological quality in the Sanjiangyuan National Park. *Sustainability* 12(16):6634. <https://doi.org/10.3390/su12166634>
- Usman K, Ramdhani M (2019) Comparison of classical interpolation methods and compressive sensing for missing data reconstruction. In: 2019 IEEE international conference on signals and systems (ICSigSys). IEEE. <https://doi.org/10.1109/icsigsys.2019.8811057>
- Wang J, Rich PM, Price KP (2003) Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. *Int J Remote Sens* 24(11):2345–2364. <https://doi.org/10.1080/0143160210154812>
- Wen Q, Gao J, Song X et al (2019) RobustSTL: a robust seasonal-trend decomposition algorithm for long time series. In: Proceedings of the AAAI conference on artificial intelligence, pp 5409–5416. <https://doi.org/10.1609/aaai.v33i01.33015409>
- Wen Q, Zhang Z, Li Y et al (2020) Fast RobustSTL: efficient and robust seasonal-trend decomposition for time series with complex patterns. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp 2203–2213. <https://doi.org/10.1145/3394486.3403271>
- Wu K, Chen J, Yang H et al (2023) Spatiotemporal variations in the sensitivity of vegetation growth to typical climate factors on the Qinghai–Tibet Plateau. *Remote Sens* 15(9):2355. <https://doi.org/10.3390/rs15092355>
- Xie Y, Chen Y, Zhang Y et al (2023) Response of vegetation normalized different vegetation index to different meteorological disaster indexes in karst region of Guangxi, China. *Heliyon*. <https://doi.org/10.1016/j.heliyon.2023.e20518>
- Yin G, Hu Z, Chen X et al (2016) Vegetation dynamics and its response to climate change in Central Asia. *J Arid Land* 8(3):375–388. <https://doi.org/10.1007/s40333-016-0043-6>
- Yoo JC, Han TH (2009) Fast normalized cross-correlation. *Circuits Syst Signal Process* 28:819–843. <https://doi.org/10.1007/s00034-009-9130-7>
- Yuan X, Li L, Chen X et al (2015) Effects of precipitation intensity and temperature on NDVI-based grass change over northern china during the period from 1982 to 2011. *Remote Sens* 7(8):10164–10183. <https://doi.org/10.3390/rs70810164>
- Zeng Y, Jia L, Menenti M et al (2023) Changes in vegetation greenness related to climatic and non-climatic factors in the Sudano-Sahelian region. *Reg Environ Change*. <https://doi.org/10.1007/s10113-023-02084-5>
- Zeren Cetin I, Varol T, Ozel HB et al (2023) The effects of climate on land use/cover: a case study in Turkey by using remote sensing data. *Environ Sci Pollut Res* 30(3):5688–5699. <https://doi.org/10.1007/s11356-022-22566-z>
- Zhang J, Roy D, Devadiga S et al (2007) Anomaly detection in MODIS land products via time series analysis. *Geo-Spat Inf Sci* 10(1):44–50. <https://doi.org/10.1007/s11806-007-0003-6>
- Zhang Z, Liang S, Xiong Y (2023) Vegetation dynamics and their response to climate changes and human activities: a case study in the Hanjiang River Basin, China. *Forests* 14(3):509. <https://doi.org/10.3390/f14030509>
- Zhao X, Tan K, Zhao S et al (2011) Changing climate affects vegetation growth in the arid region of the Northwestern China. *J Arid Environ* 75(10):946–952. <https://doi.org/10.1016/j.jaridenv.2011.05.007>
- Zhou J, Hua Z (2021) A new tendency correlation coefficient for bivariate time series. *Rendiconti Lincei Scienze Fisiche e Naturali* 32(3):479–491. <https://doi.org/10.1007/s12210-021-00992-4>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.