IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PRINCIPAL COMPONENT ANALYSIS (PCA) IN THE EVALUATION OF VEGETATION INDICES DERIVED FROM TIME-SERIES REMOTE SENSING DATA: A REVIEW

Arun.S

Guest Lecturer,
Department of Physical Oceanography,
KUFOS University, Panangad, Kochi, India.

Abstract: Big datasets encompass a large volume of information, but they can be hard to decipher. Principal components analysis (PCA) is a reliable technique in multivariate data analysis reducing the number of parameters while retaining as much variance as possible. PCA is a technique for reducing data dimensionality and enhancing interpretability while helping to reduce information loss. It executes this by generating new uncorrelated variables that optimize variance in a sequential manner. PCA is an effective method of data analysis as it reduces seeking new variables, the principal components, to resolving an eigenvalue/eigenvector, and the new variables are represented by the dataset at the time of creation. To enhance regional level variations in multi-temporal sets of data, the PCA is being used as the first method of data transform technique. Cross-validation techniques may be used to determine the PCA model's value. PCA is commonly used to detect changes in time series data and is one of the most popular techniques due to its efficiency and ability to enhance even subtle modifications It's adaptable in another way, too, because different kinds of techniques were created for different data types and structures. And, after analyzing thousands of original variables, the first few components that account for the bulk of the variance in the data are investigated. These new variables can be visualized and statistically analyzed to find similarities and discrepancies between samples.

Index Terms - Principal component analysis (PCA), Time-Series Remote Sensing, VHI, SMCI

I. INTRODUCTION

For several earth system processes, the vegetation cover is considered a significant variable (Hansen, DeFries, Townshend, & Sholberg, 2000). Sabins (1996) mentioned that natural and cultivated vegetation covers the earth and highly affects the ecosystem. Vegetation indices are key indicators illustrating the vigour of greenery. They are good at accurately sensing green vegetation as single spectral channels. (Campbell, 1987). In past years, data on the preparation, strength, and quantity of the globe's vegetation was also inadequate. Vegetation, soil light, carbon emission, darkness, soil color, and precipitation all contribute to the reflectance spectra of plant cover (Asrar et al.1984). The concept of a vegetation index is well fitted to representing vegetation over vast areas, such as regions that cover numerous pixels in a picture.. (Sabins, 1996) stated that satellite remote sensing gives the facility to regularly store the Land surface resources vegetation parameter extracted from long time-series multi-temporal NDVI datasets from the National Oceanographic Atmospheric Administration (NOAA). Advanced Very High-Resolution Radiometer (AVHRR) by using PCA may classify the natural vegetation depending on either the temporal differences of magnitude and seasonal variation of Normalized Difference Vegetation Index (NDVI) in the province of Arizona. (Yuji et al.,1996). It gathers data at varying spatiotemporal scales and provides a plethora of data sets ranging from low spatial resolution MODIS data to finer-resolution satellite data sources like Worldview -2. Seasonal vegetation patterns being identified through the MODIS NDVI time-series covering two decades, and vegetation occurrences were detected (Richards, 1986) and analyzed by PCA.

PCA was carried out using the NDVI as an input variable. Using PCA as both an analytical and cost-effective method for mining-related details in NDVI time-series research involving vegetation inter-annual anomalies. To enhance regional level variations in multi-temporal sets of data, the PCA is being used as the first method of data transform technique. (Lasaponara, 2006), (Howarth et al.2006). Geospatial analysis was used to better process the PCA results to classify and analyze the land degradation process. The monthly Maximum Value Composite (MVC) of NDVI obtained from onboard MODIS satellites are used in the study (Kundu et al., 2018), (Wim 1999). The enhanced NDVI profiles are created using the S-G method's moving window algorithm, which eliminates outliers and corrects errors in time series data. (Chen et al.,2004). Timesat software is intended to practice time-series of vegetation index resulting from satellite spectral dimensions. The series of fiscal and communal assets in rural extents have impacts on alteration in land cover patterns by climatic variability and anthropogenic actions and it eventually disturbs the urban inhabitants (Mangiarotti et al. 2010).

1.1 Time series remote sensing satellites

Sabins (1996) mentioned that satellite remote sensing enables our first opportunity to systematically repetitively store the Earth's surface resources. The benefits of satellite-based remote sensing include high spatial resolution, which allows for reliable and comparable data to be collected from long-time series at a minimal price. (Foley et al.,1998). Landsat series 7 and 8 satellite systems, for instance, give open access to visible and multispectral datasets. The complexity to use these satellites for agricultural applications is attributed to the per-pixel resolution (30 m for Landsat and 500 m for MODIS) and the orbital period (16 days for Landsat and 26 days for SPOT). The WorldView-2 spacecraft from Digital Globe is the first-ever commercial high-resolution satellite means of supplying eight visible spectral sensors as well as near-infrared sensors. Almost every sensor is specifically based on a broad electromagnetic spectrum range, in addition to the four standard multispectral bands: blue (450–510 nm), green (510–580 nm), red (630–690 nm), and NIR (770–895 nm). The estimated revisited period of 16 days has been the next major limitation of satellite-based earth observation, which makes it more difficult for agricultural applications, especially plant nutrient and water conservations. (Jinru & Baofeng, 2017).

Table 1.1 Satellite sensors and their characteristics

Primary Use	Band	Bandwidth[1]	Spectral Radiance[2]	Required SNR[3]
Land/Cloud/Aerosols Boundaries	1	620-670	21.8	128
	2	841-876	24.7	201
Land/Cloud/Aerosols Properties	3	459-479	35.3	243
	4	545-565	29.0	228
	5	1230-1250	5.4	74
	6	1628-1652	7.31	275
	7	2105-2155	1.0	110
OceanColor/ Phytoplankton/ Biogeochemistry	8	405-420	44.9	880
	9	438-448	41.9	838
	10	483-493	32.1	802
	11	526-536	27.9	754
	12	546-556	21.0	750
	13	662-672	9.5	910
	14	673-683	8.7	1087
	15	743-753	10.2	586

	16	862-877	6.2	516
Atmospheric Water Vapour	17	890-920	10.0	167
	18	931-941	3.6	57
	19	915-965	15.0	250

Primary Use	**Band**	**Bandwidth[1]**	**Spectral Radiance[2]**	**Required NE[]T(K)[4]**
Surface/Cloud Temperature	20	3.660-3.840	0.45(300K)	0.05
	21	3.929-3.989	2.38(335K)	2.00
	22	3.929-3.989	0.67(300K)	0.07
	23	4.020-4.080	0.79(300K)	0.07
Atmospheric temperature	24	4.433-4.498	0.17(250K)	0.25
	25	4.482-4.549	0.59(275K)	0.25
Cirrus clouds Water Vapour	26	1.360-1.390	6.00	150(SNR)
	27	6.535-6.895	1.16(240K)	0.25
	28	7.175-7.475	2.18(250K)	0.25
Cloud Properties	29	8.400-8.700	9.58(300K)	0.05
Ozone	30	9.580-9.880	3.69(250K)	0.25
Surface/Cloud Temperature	31	10.780-11.280	9.55(300K)	0.05
	32	11.770-12.270	8.94(300K)	0.05
Cloud Top Altitude	33	13.185-13.485	4.52(260K)	0.25
	34	13.485-13.785	3.76(250K)	0.25
	35	13.785-14.085	3.11(240K)	0.25
	36	14.085-14.385	2.08(220K)	0.35

2. PRINCIPAL COMPONENT ANALYSIS

PCA is a linear transformation that eliminates redundancy by converting or rotating the original feature space axes, so the drought information can be shown in a new model space without correlation (Lasaponara, 2006). The first principal components, which have a higher proportion of variance than in the subsequent components, and also similar anomalies or spatial distributions, are vested. As a whole, it is frequently considering only a few principal components (Antonio et al., 2015). PCA transformation produces different core components (PC1, PC2, PC 3 ... etc.), which are uncorrelated and organized according to the proportion of variance they reflect in the original band collection. Those bands are often easier to interpret than the data from the source. PCA is commonly used to detect changes in time series data and is one of the most popular techniques due to its efficiency and ability to enhance even subtle modifications. The PCA method first calculates the covariance matrix, eigenvalues, and eigenvectors for all input data then calculates the percent of overall data set variance represented by each component, and finally generates a series of new data (called Eigen channels or components) by multiplying the original input data eigenvector (Lasaponara, 2006). The matrix of covariances measures both the Eigenvector and the values. Even if the PC transform greatly increases data variance in the first few PC bands, those bands typically make up a large part of the coherent image information and are used to distinguish common data features. (Richard, 1999**).** Mysterious functions and/or noise are common in higher-order PC images. PCA techniques are widely used in dimensionality reduction and also have hyperspectral image processing applications. The spectral component of PCA transformations is often defined by a significant level of spatial correlation with images in different bands for most hyperspectral RS applications. (Pu & Gong., 2004, Chaurasia & Dadhwal., 2004, Chen & Qian., 2011, Zhou et al., 2015). Considering the large amount of spectral similarity between the spectral reflectance of different leaves, it should be simple to reconstruct leaf spectra using a small number of main components by using sample dimension PCA transformation rather than spectral dimension transformation. The PCA method offers a precise and reliable computation technique for the reconstruction of leaf spectral reflectance as well as biochemical leaf material (Liangyun et al.,2017**).** The need for PCA methods to explain the temporal and spatial variations of vegetation cover is common. PCA may classify the peak relevant modes of variability in time and space. It uses a formula to calculate a collection of orthogonal principal components, resulting in data that is completely original despite all variance (Li and Kafatos,2000). In addition, PCA rotates the spot-scatter around its center pixel and integrates with the coordinate axes to optimize the propagation of the data projected onto them. For informative reasons, authors might state to PCA as a rotation of the coordinate axes (Pielou, 1984). This spread is the sum of squares of the data's coordinate points along the axes, also known as principal component ratings. In addition, the new axes are not correlated. The scores individuated by centered data are expressed as deviations from the mean of the variable. The rotation translates mathematically in a weighted summation of the original dimensions.

S.No.	Purpose	PCA Technique	Study conducted	Formula / Software	Reference
1	Public Irrigation Schemes	Linear aggregation method & orthogonal varimax method	Kenya	$W_k = \sum_{j=1}^{j=n} \frac{(\text{Factor loading}_{kj})^2}{\text{eigenvalue}_j} \times \frac{\text{eigenv}}{\sum_{j=1}^{j=n} \text{eige}}$	Faith et al.2018
2	Leaf Reflectance Spectra & Biochemical Contents	Partial least squares regression (PLSR)	Italy	$\tilde{R}(\lambda) = \sum_{i=1}^{n} k_i \times \boldsymbol{\Phi}_i($	Liangyun et al.2017
3.	Land cover Characterisation	Correlation Matrix	Arizona		Yuji et al.1996
4.	Vegetation Characteristics	Cluster Analysis	North America	IDRISI32 Release 2	Shaw-wen, 2004
5.	Temporal analysis of rainfall and vegetation index	Factorial Analysis method	Pernambuco, Brazil	$\mathbf{U} = \mathbf{A}^t.\mathbf{X}$	Dantas et al.2016
6.	Vegetation dynamics	Spectral analysis	Zimbabwe	IDRISI Selva	Seth et al.2013
7.	Land cover change	Multi-temporal image analysis	Ghana		Nyamekye et al.2014
8.	Wetland Change Detection	Correlation Matrix	Zambia	ERDAS IMAGINE	Christopher,2004
9.	Accuracy of Classification Maps	Multi-Spectral band analysis	North of Iran	ENVI	Masoumh & Amir.2016
10.	Remote sensing	covariance matrixes	Valencia province, Spain	$Y_6 = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_6 \end{pmatrix} = \begin{pmatrix} w_{1,1} & \cdots & w_{1,6} \\ \vdots & \ddots & \vdots \\ w_{6,1} & \cdots & w_{6,6} \end{pmatrix}$	Javier et al.2013
11.	Evaluating Vegetation Anomalies	covariance matrixes	Italy	$P_i = \sum_{i=1}^{n} P_k \times u_{k,i}$	Lanorte et al.2015

Table. 2.1 PCA techniques used in remote sensing studies

2.1 Adaptations of principal component Analysis

The basic principle of PCA is obvious which leads to low-dimensional representations of complex sets of data adaptively and insightfully. The subsections also revealed several subtleties that add a layer of complexity. Besides that, PCA can be updated in a variety of ways to achieve different objectives or examine different types of data (Hotelling, 1993). Since PCA is used in many distinct aspects, there is a lot of testing into advancements and modifications from various disciplines. The four adaptations mentioned in the following subsections are functional PCA, PCA changes to simplify interpretations, RPCA, and symbolic data PCA, which were chosen at random from the many that remained. Functional PCA, PCA changes to enhance representations, RPCA, as well as symbolic data PCA are the four most common adaptations, which are briefly listed below.

2.1.1 Functional principal component analysis

Pioneering study on functional PCA (Rao, 1958) operated a standard PCA on a data matrix ($n{\times}p$) acquired by sampling n curves x (t) at each p point in time (t, j=1, ... ,p), so that when the data matrix element in row I column j is x(t). Following that, the p-dimensional loading vectors obtained from a PCA of this data matrix are interpreted as sampled main functions, which can be filtered to recover functional form and interspersed in the observed curves as primary variability (Ramsay & Silverman., 2006). And there is a correlation pattern, along with similarities between the p variables among the n observations. A 'dynamic' variant of functional PCA is recommended in (Hormann et al.,2015), which is significant for correlations among the curves identified, and for the apparent correlation between the curves.

This is focused on the concept for vector time series initially stated by (Brillinger, 1981) and use frequency domain analysis.

2.2.2 Simplified principal components

In the case of optimizing variance in q dimensions, PCA provides the strongest quality portrayal of a p-dimensional dataset in q dimensions ($q < p$). Even then, a drawback is that the new variables it describes are typically linear functions of all original variables in p. A set of PCA modifications were being indicated that attempt to construct a simpler interpretation of the q dimensions, even when minimizing loss of variance due to failure to use the PCs themselves. The idea of rotating PCs is borrowed from the factor analysis (Cattell, 1978). Rotation can facilitate faster understanding as no variation is lost when presented concerning the q-dimensional space which is rotated, as the number of variances of the q rotated components is the same as those of unrotated components. A further approach to PC simplification is the application of a constraint on the new variables loads. Often there are several versions of this approach, including the one that adapts the LASSO (the least absolute shrinkage and selection operator) linear regression approach (Jolliffe, 2003). PCA modifications under which many coefficients are completely null under specifically identified as limited versions of PCA, and research on these PCs has been extensive in past years (Hastie et al.2015). A distinction between both the methods to rotation and constraints is that it still has the benefits of interpreting those loadings exactly to zero in linear functions, while rotation is generally not involved.

2.2.3 Robust principal component analysis (RPCA)

PCA is insensitive to outliers, and even to the presence of outliers for gross data-set errors. This led investigators to distinguish robust PCA variants, and the term RPCA was coined to refer to various approaches to this problem. Robust approaches to matrices **(**Huber, l977; Huber, 1981**)** of covariance or correlation, and ways to represent robust PCs, were studied. (Wright et al.2009 expressed RPCA as a division of the X data matrix n / p into two np components. (Zhao et al.2016) explored the role of alternative algorithms on image processing and image recognition problems in the presence of different types of 'noise'. Algorithms for component analysis are discussed in (Candeset al.2011),which is an important subject despite the mathematical complexities involved.

2.2.4 Symbolic data principal component analysis

Symbolic data refers to more complex data structures, such as intervals or histograms. (Bock et al.2000; Brito, 2014). Interval data appears when it is necessary to keep track of an observation's internal variability.PCA modifications for most of these data (Ichino & Yaguchi, 1994) compute PCs that are often in the form of an interval and may also display ranges of values**.**Histograms, which can be interpreted as a conceptualization of interval-valued data, with multiple intervals and corresponding frequencies for each measurement, are perhaps common method of symbolic research. A recent analysis outlines several (Makosso, 2015) proposed principles for histogram data from PCA studies.

2.2.5 Standardization and correlation matrix

PCs is initially believed to be a good estimate of the centered variables. The PCA variables also have some drawbacks because they use different measurement units. To resolve this undesirable feature, it is normal practice to standardize the variables. All data value of *xij* is centered and also divided by the standard deviation *sj* of the vector n observations of variable *j* (Jolliffe & Jorge., 2016).

$$zij = \frac{xij - xj}{sj}$$

The standardized data matrix Z, whose jth column is vector zj with the standardized n observations of variable j, replaces the current data matrix X. Since a standardized dataset's covariance matrix is literally the original dataset's correlation matrix R, the standardized data in a PCA is often known as a correlation matrix PCA.. In contrast, certain statistical software assumes that a PCA requires a PCA correlation matrix and, in a few situations, the normalization used for loading vectors a_k of correlation matrix PCs is not $a_k a_k = 1$. The coefficient of correlation between the vector jth and the PC kth are set by in a matrix of correlation PCA. (Jolliffe, 2002).

$$^{r}\mathrm{var}_{j,}\mathrm{PC}_k = \lambda_k a_{jk.}$$

As a result, instead of using $a_k a_k = \lambda_k$ the normalisation of $a_k a = 1$ is often used, and the coefficients of the new loading vectors ak are the correlations in between original variable and kth PC.

3. CONDITIONED INDICES AND VALIDATION PROCESS

Onboard the TERRA satellites, NASA operated a Moderate Resolution Imaging Spectro-radiometer (MODIS) sensor with a sweeping swath of 2330 km and 1-2day temporal resolution in 36 multi-spectral bands, swiftly extending earth observation with a higher resolution Series of NDVI data (http://terra.nasa.gov/). The NDVI was first developed as a tool for assessing the health and density of vegetation. (Tucker 1979) .The NDVI is computed with the formula

$$NDVI = (NIR - red / NIR + red) \qquad (1)$$

where, NIR and red are the reflectances in the NIR and red bands, respectively. Plant vigor, green cover percentage, Leaf Area Index (LAI), and biomass content are all measured by the NDVI. While used to monitor vegetation cover on the earth's surface, the first two spectral bands of satellite data, Red (0.58 to 0.68 m) and Near-Infrared (0.75 to 1.1 m) are found to be useful for mapping clouds and land surfaces and illustrating surface water bodies, respectively. (Tucker et al., 2004, 2005). The NDVI value varies between -1 and +1. The presence of dense green leaves is highly likely if the NDVI value is close to +1. When the NDVI is close to zero, green leaves are absent, and the region may be densely populated. Healthy vegetation (chlorophyll) reflects more Near-Infrared (NIR) and green light than other wavelengths. It also holds the lighter red and blue. As a result, our eyes perceive plants as being green in color. The plants will be abundant if you could see near-infrared. Global vegetation dynamic trends studies using NDVI data have been perfectly-acknowledged (Tucker et al., 1985, 1986, 1991). The amount of photosynthetically active radiation (400-700 rim) absorbed by green vegetation is largely associated with the NDVI. (Asrar et al., 1984).Its logical forecaster of biomass production (Tucker, 1980; Prince, 1991)and carbon fixation(Fung et al., 1987) is its time elementary.

In detail, it should be able to relentlessly expose by radiative transfer theory that NDVI reveals the health and movement of leaf chlorophyll pigments (Myneni et al., 1995). The NDVI is globally used as a remote sensing vegetation index due to its consistency in detecting vegetation dynamics and its ease of computation and analysis. That is the foremost reason for its broader use and larger acceptance compared to another satellite-based spectral vegetation index. The use of vegetation index specifies a considerable greening trend from cumulative seasonal amplitude and growing season spell in the northern high latitudes landscapes**.** (Myneni et al., 1998). The reason for seasonal phasing is to base a classification nonappearance of the spatial NDVI profile comparatively than on the NDVI values themselves (Lloyd 1990; Samson 1993). Such appearances consist of overall 'greenness', measured by the combined NDVI all over the year beyond a threshold value; seasonality, measured by the amplitude of the NDVI profile and speediness of brown off and also measured by the maximum negative slope in the profile. The NDVI is a spectral vegetation index that depicts the variation in near-infrared and red reflectance bands defined by their sum. (Curran and

Steven 1983). The greater the NDVI value indicates more the concentration of green vegetation. The related studies of using NDVI have discovered a close association between NDVI and seasonal rainfall in semi-arid regions (Tucker and Nicholson 1999; Anyamba and Eastman 1996). PC1 describes the most variance in NDVI integrals and reflects the typical built-in NDVI pattern in space. PC1 shows the standard NDVI over the entire array. PC2 describes the overall remaining variance not described by PC1, and follows the same logic. Resultantly, the components after PC1 (particularly PC2) tend to represent specific events such as fires, drought periods, and, in specific, vegetation-related land degradation phenomena, instead of showing the general development over the period.

The normalized vegetation difference index (NDVI)-based vegetation condition index (VCI) (Kogan and Sullivan, 1993; Kogan, 199) and Temperature condition index (TCI) (Kogan, 1995) are two effective tools for mapping the severity, extent, and impacts of drought at a regional or global level (Singh et al.2003). Earlier studies indicate that VCI is ideal for measuring large-scale impacts of drought on vegetation, including agricultural drought, and that VCI has a strong correlation with crop yields (Liu and Kogan, 1996; Unganai and Kogan, 1998; Kogan et al., 2005; Salazar et al., 2007, 2008). Research in semi-arid areas of the Iberian Peninsula also noticed that the assessment of VCI was more challenging than other indices of drought because it provides an indirect measure of conditions of moisture (drought). VCI can represent everything that stresses the vegetation namely pests, diseases, and nutrient deficits (Vicente-Serrano, 2007). In addition, some studies observed that the VCI and TCI analyzing together will be better than individually, and formed a vegetation health index (VHI) (Kogan, 1997; Kogan et al., 2004). Moreover, in high-latitude humid areas, where vegetation growth is predominantly restricted by lower temperatures in contrast to low-latitude regions, care has to be taken to monitor drought with VHI (Karnieli et al., 2006). Not all these indices include climate data, such as precipitation and variability in soil moisture, which is among the factors that caused drought in semi-arid regions.

Table 3 1. Equations of conditioned indices used for PCA studies

Indices	**Formula**	**Source**
Vegetation Condition Index (VCI)	$\frac{(NDVI_j - NDVI_{min})}{(NDVI_{max} - NDVI_{min})}$	(Kogan, 1995)
Temperature Condition Index (TCI)	$\frac{(LST_{max} - LST_j)}{(LST_{max} - LST_{min})}$	(Du et al., 2013)
Soil Moisture Condition Index (SMCI)	$\frac{(SM_j - SM_{min})}{(SM_{max} - SM_{min})}$	(Hao et al., 2015)
Precipitation Condition Index (PCI)	$\frac{(PCI_j - PCI_{min})}{(PCI_{max} - NDVI_{min})}$	(Jiao et al., 2019)
Vegetation Health Index (VHI)	VHI $a * \text{VCI} + (1 - a) * \text{TCI}$	(Kogan, 1995)

The MODIS Terra data source can be used to measure VCI, TCI, and PCI, and SCI. The soil moisture index (SCI) has also been specified to detect the moisture in the soil deficits which affect crop stress and crop yield. The key component incorporates VCI, TCI, and PCI to describe Integrated Drought Severity Index (IDSI) that can detect drought occurrence, duration, and intensity (Jeganathan et al.,2015). The IDSI utilizes precipitation deficits, soil stress conditions, and plant growth progress in the cycle of drought, and it helps in observing severe drought. For instance, this method monitored the time series drought from 2000 to 2014 to evaluate the output in selected locations, including Sri Lanka, and Southern Indian states, and its findings were consistent with the actual drought conditions in recent decades. (IWMI, 2005).

4. CONCLUSIONS

The emission spectrum among vegetated areas poses a complex combination of vegetation, brightness of soil, environmental impacts, shadow, the color of the soil, and humidity. The concept of vegetation index was well apt for vegetation over massive areas, eg. Over regions that cover several pixels of an image. PCA-allied approaches are often provided with a substantial role in added statistical processes, like linear regression (with major component regression (Jolliffe, 2002) and smoothened the sequential clustering of individuals and variables (Vichi & Saporta, 2009). While methods like correspondence analysis, canonical correlation analysis, and linear discriminant analysis are only interrelated to PCA, they share a common methodology in that they are focused on factorial decompositions of some matrices. PCA can be used as a mapping tool that shows substantial interannual variations and unidirectional variations in vegetation studies. This adaptive analysis technique is among the oldest and broadly used due to its dimensionality reduction functionality making it easier to interpret complex data types for the user community. The satellite-based conditioned indices such as VCI, TCI, PCI, and SMCI can monitor the significance of vegetation growth and crop stress, duration, extent, and severity of drought-related conditions.

REFERENCES

1. Antonio Lanorte, Teresa Manzi, Gabriele Nole, and Rosa Lasaponara., (2015). On the Use of Principal Component Analysis (PCA) for evaluating the Seasonal Vegetation anomalies from LANDSAT-TM NDVI Temporal series in the Basilicata Region (Italy), ICCSA, 204-216.
2. Anyamba, A., and Eastman, J. R.,(1996). Interannual Variability of NDVI over Africa and Its Relation to El Nino/Southern Oscillation. *International Journal of Remote Sensing* 17: 2533–2548.
3. Asrar, G., Fuchs, M., Kanemasu, E. T, and J. L. Hatfield., (1984). Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat, *Agron. j.,* 76, 300-306.
4. Bock, H.H., Diday, E., (2000). *Analysis of symbolic data*. Berlin, Germany: Springer.
5. Brillinger, D, R., (1981). Time series: data analysis and theory, Expanded edn., CA: Holden-Day.
6. Brito, P., (2014). Symbolic data analysis: another look at the interaction of data mining and statistics. *WIREs Data Mining Knowl. Discov*. 4, 281–295. (doi:10.1002/widm.1133).
7. Campbell, J. B. (1987) *Introduction to Remote Sensing.* The Guilford Press, New York, USA, 551.
8. Campbell, J. B. (2002).) *Introduction to Remote Sensing* (3rd ed). New York: The Guilford Press.
9. Campos-Taberner, M., Moreno-Martinez, A., García-Haro, F.J., Camps-Valls, G., Robinson, N.P., Kattge, J., Running, S.W., (2018). Global Estimation of Biophysical Variables from Google Earth Engine Platform. *Remote Sens.*, *10*, 1167.
10. Candes, E.J., Li, X., Ma, Y., Wright, J., (2011). Robust principal component analysis. *J. ACM* 58.
11. Cattell, R. B., (1978). The scientific use of factor analysis in behavioral and life sciences., NY: Plenum Press.
12. Chaurasia, S.; Dadhwal, V. K., (2004). Comparison of principal component inversion with VI-empirical approachfor LAI estimation using simulated reflectance data. *Int. J. Remote Sens*., *25*, 2881–2887.
13. Chen, G.; Qian, S.E. (2011). Denoising of hyperspectral imagery using principal component analysis and waveletshrinkage. *IEEE Trans. Geosci. Remote Sens*. *49*, 973–980.
14. Chena Jin, Per. Jonssonc, Masayuki Tamurab, Zhihui Gua, Bunkei Matsushitab, Lars Eklundhd., (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Elsevier, *Remote Sensing of Environment* 91 .332 – 344.
15. Curran, P. J., and M. D. Steven., (1983). Multispectral Remote Sensing for the Estimation of Green Leaf Area Index. *Philosophical Transactions of the Royal Society of London, Mathematical and Physical Sciences* 309: 257–270.
16. DeFries, R. S, Hansen, M.C., & Townshend, J. R. G. (2000). Global continuous fields of vegetation characteristics: a linear mixture model applied to multi-year 8 km AVHRR data. International Journal of Remote Sensing, 21, 1389-1414.
17. DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., & Loveland, T. R. (2000). A new global 1km data set of percent tree cover derived from remote sensing. Global Change Biology, 6, 247-254.

18. DeFries, R.S., &Belwad, A.S. (2000). Global and regional land cover characterization from satellite data: an introduction to the Special Issue. International Journal of Remote Sensing, 21, 1083-1092.

19. Du, L.; Tian, Q.; Yu, T.; Meng, Q.; Jancso, T.; Udvardy, P.; Huang, Y. (2013). A comprehensive drought monitoring method integrating MODIS and TRMM data. *Int. J. Appl. Earth Obs.*, *23*, 245–253, doi: 10.1016/j.jag.2012.09.010.

20. Eklundh and Johsson., (2012). Timesat 3.1 software manual, Lund University, Sweden.

21. Fung, I. Y., C. J. Tucker, and K. C. Prentice., (1987). Application of advanced very high resolution radiometer vegetation index to study atmosphere- biosphere change of CO2,*J. Geophys R.es.*,9 2, 2999-3015.

22. Hansen, M. C., DeFries, R. S., Townshend, J. R. G., & Sholberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing, 21, 1331-1364.

23. Hao, C.; Zhang, J.; Yao, F. **(**2015**).** Combination of multi-sensor remote sensing data for drought monitoring over Southwest China. *Int. J. Appl. Earth Obs.*, *35*, 270–283, doi:10.1016/j.jag.2014.09.011.

24. Hastie, T., Tibshirani, R., Wainwright, M., (2015) . *Statistical learning with sparsity: the LASSO and generalizations*. Boca Raton, FL: CRC Press

25. Hormann, S., Kidzinski, L., Hallin, M., (2015). Dynamic functional principal components. *J. R. Stat. Soc. B* **77**, 319–348. (doi:10.1111/rssb.12076).

26. Hotelling H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417–441, 498–520. (doi:10.1037/h0071325).

27. Howarth, P., Piwowar, J., Millward, A., (2006). Time-Series Analysis of Medium-Resolution, Multisensor Satellite Data for Identifying Landscape Change. *Photogrammetric Engineering and Remote Sensing*72(6), 653–663

28.Huber, P.J. (1977) *Robust statistical procedures*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

29. Huber, P.J. (1981). *Robust statistics*., NY, Wiley.

30. Ichino, M., Yaguchi, H., (1994). Generalized Minkowski metrics for mixed feature type dataanalysis. *IEEE Trans. Syst. Man Cybern.* **24**, 698–708. (doi:10.1109/21.286391).

31. Jiao, W.; Wang, L.; Novick, K.A.; Chang, Q. **(**2019**).** A new station-enabled multi-sensor integrated index for drought monitoring. *J. Hydrol.*, *574*, 169–180, doi:10.1016/j.jhydrol.2019.04.037.

32. JinruXue., and Baofeng Su, (2017), Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications., Hindawi, *Journal of Sensors*,https://doi.org/10.1155/2017/1353691.

33. Jolliffe I.T, Trendafilov, N., Uddin, M., (2003). A modified principal component technique based on the LASSO. *J. Comput. Graph. Stat.* **12**, 531–547. (doi:10.1198/1061860032148)

34. Jolliffe, I.T., (2002). Principal component analysis, 2nd edn. New York, NY: *Springer-Verlag.*

35. Karnieli, A., Bayasgalan, M., Bayarjargal, Y., Agam, N., Khudulmur, S., Tucker, C.J., (2006). Comments on the use of the vegetation health index over Mongolia. *International Journal of Remote Sensing* 27 (10),

36. Kasturirangan, K., (1996). Ground based measurements for interpretation of vegetation indices over developing countries, *Advanced Space Research*, 17(8), pp 31-40

37. Kogan, F., Roytman, L., (2008). Using vegetation health indices and partial least squares method for estimation of corn yield. *International Journal of Remote Sensing*, 29 (1), 175–189.

38. Kogan, F.N. (1995). Application of vegetation index and brightness temperature for drought detection. *Adv. Space Res.*, *15*, 91–100, doi:10.1016/0273-1177(95)00079-t.

39. Kogan, F.N., (1995). Droughts of the late 1980s in the United States as derived from NOAA polar orbiting satellite data. Bulletin of the American Meteorological Society 76 (5), 655–668.

40. Kogan, F.N., (1997). Global drought watch from space. Bulletin of the American Meteorological Society 78 (4), 621–636.

41. Kogan, F.N., Yang, B., Wei, G., Zhiyuan, P., Xianfeng, J., (2005). Modelling corn production in China using AVHRR-based vegetation health indices. International Journal of Remote Sensing 26 (11), 2325–2336.

42. Kundu, A., Dwivedi, S. & Dutta, D., (2016). Monitoring the vegetation health over India duringcontrasting monsoon years using satellite remote sensing indices. *Arabian Journal of Geosciences*, 9:144-159.

43. Lasaponara, R., (2006). On the use of principal component analysis (PCA) for evaluating Interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series. *Ecol. Model*. 194, 429–434.

44. Li, Z., & Kafatos, M., (2000). “Interannual Variability of Vegetation in the United States and Its Relation to El Niño/Southern Oscillation.” *Remote Sensing of Environment* 71: 239–247.

45. LiangyunLiu., Bowen Song., Su Zhang., and Xinjie Liu. (2017). A Novel Principal Component Analysis Method for the Reconstruction of Leaf Reflectance Spectra and Retrieval of Leaf Biochemical Contents. *Remote Sens*. 9, 1113; doi:10.3390/rs9111113

46. Lloyd, D., (1990) A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. *International Journal of Remote Sensing*, 11, 2269-2279.

47. MakossoKallyth, S., (2015). In press. Principal axes analysis of symbolic histogram variables. *Stat. Anal. Data Mining*. doi:10.1002/sam.11270.

48. Mangiarotti, S., Mazzega, P., Hiernaux, P., & Mougi, E. 2010. “The Vegetation Cycle in West Africa from AVHRR–NDVI Data: Horizons of Predictability Versus Spatial Scales.” *Remote Sensing ofEnvironment* 114: 2036–2047.

49. Myneni, R. B., Hall, F.G., Piers, J. Sellers., and Marshak, A. L., (1995). The Interpretation of Spectral Vegetation Indexes. IEEE, *Geoscience and Remote Sensing*,33,2.

50. Myneni, R. B., Tucker, C. J., Asrar. G., Keeling, C. D., (1998). Interannual variations in satellite-sensed vegetation index data from 1981 to 1991. *Journal of Geophysical Research*, 103, NO. D6.

51. Pielou, E.C., (1984). The Interpretation of Ecological Data: a Primer on Classification and Ordination, Hardcover edn. New York: Wiley.

52. Press, H., Flannery, B., Teukolsky, A., Vetterling, W., (1992). Numerical Recipes in C, Cambridge 53. University Press, New York.

54. Project Report, 2015. Development of South Asia Drought Monitoring System/ International Water Management Institute (IWMI), Colombo, Sri Lanka.

55. Pu, R.L.; Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closuremapping. *Remote Sens. Environ., 91*, 212–224.

56. Ramsay, J.O., Silverman, B.W., (2006). Functional data analysis, 2nd edn. *Springer Seriesin Statistics*. NY: Springer.

57. Rao CR. (1958). Some statistical methods for comparison of growth curves. *Biometrics* **14**, 1–17. (doi:10.2307/2527726).

58. Richards, J. A., (1999.) Remote Sensing Digital Image Analysis*:* An Introduction, *Springer-Verlag*.

59. Richards, J.A., (1986). Remote sensing Digital Image Analysis. *Springer*–Verlag, Kensington, pp. 127–138.

60. Sabins, F.F. (1996). Remote sensing: principles and interpretation (3rd ed). New York: W.H. Freeman and Company.

61. Salazar, L., Kogan, F., Roytman, L., (2007). Use of remote sensing data for estimation of winter wheat yield in the United States. *International Journal of Remote Sensing*, 28 (17), 3795–3811.

62. Salazar, L., Singh, R.P., Roy, S., Kogan, F., (2003). Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India. , 24 (22), 4393–4402.

63. SAMSON, S. A., (1993). Two indices to characterize temporal patterns in the spectral response of vegetation. *Photoqrammetric Engineering and Remote Sensing*, 59, 511-517.

64. Seth, M., Kazadi Sanga-Ngoie, and Shoko, K., (2013). Vegetation dynamics of Zimbabwe investigated using NOAA-AVHRR NDVI from 1982 to 2006: a principal component analysis. *International Journal of Remote Sensing*, Vol. 34, No. 19, 6764–6779, http://dx.doi.org/10.1080/01431161.2013.806833

65. Tucke,r C. J., Pinzon J E and Brown M E., (2004). Global Inventory Modelling and Mapping Studies, Global Land Cover Facility, University of Maryland, College Park, Maryland, 1982-2006.

66. Tucker, C. J., Pinzon J E., Brown M E., Slayback D., Pak E W., Mahoney R., Vermote E, and Saleous N El., (2005). An extended AVHRR 8-km NDVI Dataset Compatible with MODIS and SPOT vegetation NDVI Data. *International Journal of Remote Sensing*, 26(20): 4485-4498.

67. Tucker, C. J., (1980). A critical review of remote sensing and other methods for non-destructive estimation of standing crop biomass, *Grass Forage Sci*., 35, 177-182,

68. Tucker, C. J., (1996). History of the use of AVHRR data for land applications, in Advances in the Use of NOAA AVHRR Data for Land Applications, edited by G. D'Souza, *Euopean Economic Community*, pp. 1-19,

69. Tucker, C. J., and Nicholson, S. E., (1999). Variations in the Size of the Sahara Desert from1980–1997.*Ambio* 28: 587–59.

70. Tucker, C. J., Dregne, H. E., and Newcomb, W. W., (1991). Expansion and contraction of the Saharan desert from 1980 to 1990, *Science*, 253,299-301

71. Tucker, C. J., Fung, I. Y., Keeling, C. D., and Gammon, R. H. (1986). Relationship between atmospheric CO2 variations and a satellite-derived vegetation index, *Nature*, 319, 195-199.

72. Tucker, C. J., Newcomb, W. W., and Dregne, H. E., (1994). AVHRR data sets for determination of desert spatial extent, *International Journal of Remote Sensing*, 15,3547-3565.

73. Tucker, C. J., Townshend, J. R. G. and Goff, T. E., (1985). African land-cover classification using satellite data, *Science2*, 27, 369-375.

74. Verbesselt Jan., Per Jonsson., Stefaan Lhermitte., Jan van Aardt., and Pol Coppin., (2006). Evaluating Satellite and Climate Data-Derived Indices as Fire Risk Indicators in Savanna Ecosystems, IEEE, *Geoscience and Remote Sensing*, 44, 6.

75. Vicente-Serrano, S.M., (2007). Evaluating the impact of drought using remote sensing in a Mediterranean, semi-arid region. Natural Hazards 40 (1), 173–208.

76. Vichi, M., Saporta, G., (2009). Clustering and disjoint principal component analysis. *Computational statistics and Data Analysis*, 53, 3194–3208. (doi: 10.1016/j.csda.2008.05.028).

77. W. J. Foley, A. McIlwee, I. Lawler, L. Aragones, A. P. Woolnough, and N. Berding, (1998). Ecological applications of near infrared reflectance spectroscopy - A tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance," *Oecologia*, 293–305.

78. Wright, J., Peng, Y., Ma, Y., Ganesh, A., Rao, S., (2009). Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization. In *Proc. of Neural Information Processing Systems* (*NIPS 2009*), *Vancouver*

79. Zhao, Q., Meng, D., Xu, Z., Zuo, W., Zhang, L., (2014). Robust principal component analysis with complex noise. In *Proc. of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June2014*. See http://jmlr.org/proceedings/papers/v32/zhao14.pdf.

80. Zhou, L., Ma, W., Zhang, H., Li, L., Tang, L. (2015). Developing a PCA–ANN Model for Predicting Chlorophyll a Concentration from Field Hyperspectral Measurements in Dianshan Lake, China. *Exposure Health*, *7*,591–602.