

Application of Bayesian networks to mapping Flood-based Farming Systems.

Issoufou Liman^{1,2,*}

Cory Whitney^{1,3}

*James Kungu*²

Eike Luedeling^{1,3,4}

May 28, 2019

Abstract

Accurate information on actual areas under given cropping/farming systems is an important input for applied agricultural research and policies. Therein, remote sensing provides a useful tool for analysing spatial metrics in an explicit manner. In this paper, we captured various characteristics of flood-based farming systems (FBFS) with remote sensing and applied a probabilistic framework for system analysis. FBFS were mapped across the Kisumu County in Kenya by incorporating uncertainties into mapping overlays. A Bayesian network (BNs) was used to describe and interlink various features of FBFS using expert knowledge. 3 years (2014-2016) of high-level MODIS VI products data were acquired in the forms of normalized difference spectral indices (NDSI). These were used as proxies for estimating various metrics of FBFS which were later used to feed the spatial data nodes of the BNs and were propagated along with the rest of other non-spatial nodes derived from expert judgement. We demonstrate that spatial explicit information can be derived from remote sensing data as fuzzy linguistic quantifiers which are suitable for representing node states in BNs. When such metrics are available, BNs are useful tools for incorporating uncertainties when mapping complex systems in a context of limited and uncertain information. The causal probabilistic reasoning embedded in the approach was tested on the Tigray region in Ethiopia; an area with totally different settings; and seems to perform incredibly well. Results were transparently generated in forms of intermediary prior spatial maps for specific metrics to ultimately be aggregated into final posterior maps of plausible areas for flood-based farming along with their spatial explicit uncertainties.

Contents

I. Introduction	3
II. Material and methods	3
II.1. Making sense of data in the context	3
II.1.1. Phenological overview of the landscape texture	3
II.1.2. Spatio-temporal patterns	3

II.2. Derivation of spatial data nodes	3
II.2.1. General Procedure	3
II.2.2. Specific procedure	4
Soil water content	4
Exposure to wetness	4
Sensitivity to flood	5
Vegetation sensitivity to water variation	5
Power of tools	5
Slope and Flow accumulation	6
Water presence	7
Temporal variation in vegetation	7
Flood at the beginning of the rainy season	8
III.Results and discussion	9
III.1. Comparative profiles of the NDSIs across different surface features	9
III.1.1. Comparative behavior of the NDSIs across biomes	9
III.1.2. Comparative distribution of the NDSIs values across different surfaces	11
III.2. Phenological overview of the landscape texture	12
III.3. Spatio-temporal variability in water and vegetation	12
References	19
¹ World Agroforestry Centre (ICRAF), United Nations Avenue, Gigiri, P. O. Box 30677-00100, Nairobi, Kenya	
² Kenyatta University, Department of Environmental Sciences	
³ Center for Development research (ZEF), University of Bonn, Genscherallee 3, D-53113, Bonn, Germany	
⁴ University of Bonn, Department of Horticultural Sciences, Auf dem Hügel 6, D-53121, Bonn, Germany	

* Correspondence: Issoufou Liman <issoufoul@gmail.com>

I. Introduction

II. Material and methods

II.1. Making sense of data in the context

II.1.1. Phenological overview of the landscape texture

We sampled approximatively 40% of pixels considering a regular grid sampling using the `spSample` function from the `sp` package.^{1,2} These time series of NDSIs were examined using boxplots. The NDSIs considered include the Normalized Difference Vegetation index,^{3,4} the Normalized Difference Flood index,⁵ the Goa's Normalized Difference Water index,^{6,7} and the mcfetters' NDWI.^{7,8} These NDSIs were chosen due to their suitability in describing water and vegetation.^{4–8}

II.1.2. Spatio-temporal patterns

The boxplots were supported with further principal components analysis (PCA)^{9,10} using a customized version of the `rasterPCA` function from the `RStoolbox` package¹¹ making used of the `prcomp` function from the `base` package¹² and other low-level functions from the `raster` package.¹³ The PCA, used to interpret the large dataset by creating a smaller number of components, was followed by other probabilistic estimations (see section II.2.) defining the spatial representations of the spatial data node states to be used in the BNs. These supporting analyses focus only on the NDVI and NDFI but a comparative analysis connecting all the NDSIs across the main land uses is provided. The outputs were presented on the map (see Supplementary materials) to give insights on how the states of vegetation and water are spatially distributed. PCA allowed us to meaningfully interpret the large dataset by creating a smaller number of components.

II.2. Derivation of spatial data nodes

II.2.1. General Procedure

The general procedure, described in more details in the main paper material, uses the different boxplot ranges (instead of thresholds) to approach the problem using imprecise quantifiers where fuzzy logic can be used in forms of Likert-type scales 55 describing the scores of the spectral responses along a range; supporting the use of BNs. These ranges were used to group the data and estimate the probability of each pixel falling within each of these ranges (Figure 4). The procedure was applied to both time series and single layer data.

II.2.2. Specific procedure

Soil water content

The node soil water content, as it says, estimates the probability of water in soil and is therefore related to soil water holding capacity. This was computed as a composite index by aggregating the **NDII6**,¹⁴ the **NDII7**,¹⁴ the **NDVI**,⁴ and the **NDFI**.⁵ While the **NDII6** and the **NDII7** perform well in detecting a mixture in different compartments (e.g. plant water, soil water, open water etc.) due to the presence of NIR and SWIR bands, the **NDFI** gives an estimate of flood conditions whereas the **NDVI** assesses the presence and condition of vegetation. Combining these in one composite index may provide a powerful tool for ecological studies since one index can be extracted from another to provide a new metric.⁵ For example, knowing that the **NDII6** broadly assess soil and vegetation water content, one could extract the **NDVI** to estimate the soil water content. Likewise, the **NDFI** can be extracted to provide an estimate of vegetation water content. We used such assumption to estimate the soil water content using the Equation (1). Such tricks where the original meaning of an index is altered by subtracting another index is commonly used in remote sensing and is by far the back bone of the spectral index theory.

$$SWC_i = \frac{1}{2} \times \left(\sum_{i=1}^n NDII6_i + \sum_{i=1}^n NDII7_i + 2 \times \sum_{i=1}^n NDFI_i - 2 \times \sum_{i=1}^n NDVI_i \right) \quad (1)$$

Where SWC_i is the soil water content, and n the total length of the time series.

For each of these 4 indices, we first computed the total over the time series. This is meant to estimate the total value scored in whatever the index measures. For example, we assumed that the sum of the **NDFI** values over the period of the 3 years considered in this study at a given pixel gives an estimate of the total water received by that pixel. After computing these total values scored, we then proceeded by summing-up the total value scored by the **NDII6** and the **NDII7** from which we extracted twice the total value scored by the **NDFI**. From the resulting value, we then extracted twice the total value scored by the **NDVI** (Equation (1)). The idea is to reduce the influence of vegetation water from a total of water contained in soil and vegetation. So doing, the remaining water would mainly be due to the above and below-ground soil water. These estimates were improved by adding to each the scored value of the **NDFI** which is another way to estimate soil wetness. We used both the **NDII6**, the **NDII7** and the **NDFI** because to take the advantage of the potential of the SWIR spectral domain as much as possible. This, however, is equivalent to doubling the real overall potential of the water content. To account for this effect, half this value was considered after reducing the amount of water stored in vegetation which was done by subtracting twice the **NDVI** to account for the influence of this vegetation water in both the **NDII6** and the **NDII7** (Equation (1)).

Exposure to wetness

This node estimates the expected total water for each cell. It was computed simply as the sum of the **NDFI** over the time series for each pixel. In this regard, it expresses the degree

to which each pixel is relatively exposed to surface water wetness. It is slightly different from the soil water content in the sense that it does not account for vegetation. Literally, this is mostly the surface soil water availability / potential regardless of whether it is used or not by vegetation. We postulate that a certain amount of water may be detected at a given pixel at some point in time without being sustained since it can runoff throughout porous soils shortly or stagnate nearly permanently under saturated or impermeable soils. Contrastingly, under soils with good water holding capacity, a relatively small amount of water can be sustained for a relatively longer period. This implies that the node exposure to wetness is more concerned about the hydrology than the actual agronomy although water exposure can be a good indicator of vegetation. It is, nonetheless, important to keep in mind that vegetation is almost always incompatible with permanently water-logged areas.

Sensitivity to flood

The node sensitivity to flood is computed as the absolute standard error of the [NDFI](#) to estimate somewhat how fast the flood index is changing throughout the time series. This node describes the relationship between water and soil. Supposing that all soil surface water originates from rainfall, and considering that this rainfall is relatively locally constant, then the local spatial variability in wetness can be explained by the relationship between soil (including land cover and topography) and water. For example, from remote sensing point of view, dry but flood-prone soils are likely to be more sensitive to torrential rainfall event compared to permanently saturated and water-logged soils.

Vegetation sensitivity to water variation

This node was also estimated as a composite index taking into account both the [NDFI](#) and the [NDVI](#). Flood followed by a rapid increase in vegetation has been used as indicator for mapping FBFS-like areas such as flooded rice.⁵ The node describes the relationship between water and vegetation using the ratio between these 2 metrics. For a given pixel, the idea conveyed by this node is simply the proportion of vegetation scored relative to amount of water received. Thus, it also implicitly gives an idea of the soil conditions because a pair of low value [NDVI](#) (absence of vegetation) and high value [NDFI](#) (water presence) may imply the presence of encrusted or swampy soils where vegetation may be sensitive to water variation.

Power of tools

The power of tools was estimated from the generic product quality assessment MODLAND-wide QA bands for each pixel and date in the time series. The MODLAND-wide QA provides a general and consistent assessment of the quality (usability and usefulness) of the MODIS products to inform the user on the quality, hence, the extent to which the results of any analysis is to be appreciated¹⁵. Many artefacts (e.g. aerosol loading, cloud, data processing algorithms, sensors failure, view angle, etc.) can introduce errors or uncertainties in remote sensing data, and it is therefore advised to consider such aspects prior to their use for subsequent analysis. While the power of tools can be estimated using band specific quality

Table 1: Translation of 2-bit pixel level-QA in MODLAND.¹⁵

2-bit encoded per pixel QA code	Decimal value	Quality attribute meanings
00	0	Pixel produced, good quality, not necessary to examine more detailed QA
01	1	Pixel produced, unreliable or unquantifiable quality, recommend examination of more detailed QA
10	2	Pixel not produced due to cloud effects
11	3	Pixel not produced primarily due to reasons other than cloud

assessment, we choose to use the MODLAND-wide QA which includes most of the artefacts in the assessment.

In the MODLAND-wide QA framework, four quality states are possible (Table 1) and the power of tools was estimated as the probability of being wrong accordingly. For each pixel throughout the time series, we remained sceptical to values greater than zero in Table 1 and the probability of being wrong was computed by dividing the count of these values by the total length of the time series. The resulting raster layer was then discretised to provide the node to be used in the BNs following the single layer procedure described in main paper material. It is worth noting that most of the poor qualities are concentrated around water bodies, hence the node power of tools was used for dual purpose in the causal model despite the little spatial variability.

Slope and Flow accumulation

Many approaches and algorithms for deriving hydrologic information from topography have been proposed in diverse conceptual forms with each having its strengths and limitations;^{16–18} mostly relying on what is known as the 3 steps conditioning procedures.^{16,17} DEM conditionings are commonly used as prerequisites for computing various topographic structures such as upslope areas, specific catchment areas, girded networks or flow paths and are extensively discussed in the literature.^{16–19} These are the filling of depressions in the original DEM, the derivation of flow directions and the computation of flow accumulation. In this paper, these were derived using [TauDEM software version 5.3](#)^{18,20–22} from within [R](#). Routines were sent via system calls from [R](#) to [TauDEM](#) and throughputs are routed from [TauDEM](#) to [R](#) to be further processed to discrete raster data following the single layer procedure described in the main paper material. [TauDEM](#) routines along with the conceptual and algorithmic details on the DEM conditioning are beyond the scope of this paper. The reader is referred to [TauDEM website](#) and related literature.^{16–18,20,21,24}

The original void-filled SRTM DEM (section Sensor Characteristics in the main paper material) was used in the 3 steps normative conditioning procedure¹⁷ to compute the slope and flow accumulation layers to account for the corresponding spatial data nodes in the BNs (see main paper material). Our approach fully accommodates for simple and looping depressions and flat areas as suggested by Jenson & Domingue¹⁷ and Tarboton.²⁰ We computed the slope as the drop distance (tangent of the slope angle; see [TauDEM website](#)) from which we derived the D-infinity flow direction as the direction outwards water flows from that pixel based on the Single-flow-direction concept.^{16,18} Based on the derived flow directions, the flow accumulation at each pixel was then calculated as the number of pixels draining into it.

Kisumu County, Kenya

MODIS NDVI: 2014-01-01 / 2016-12-26

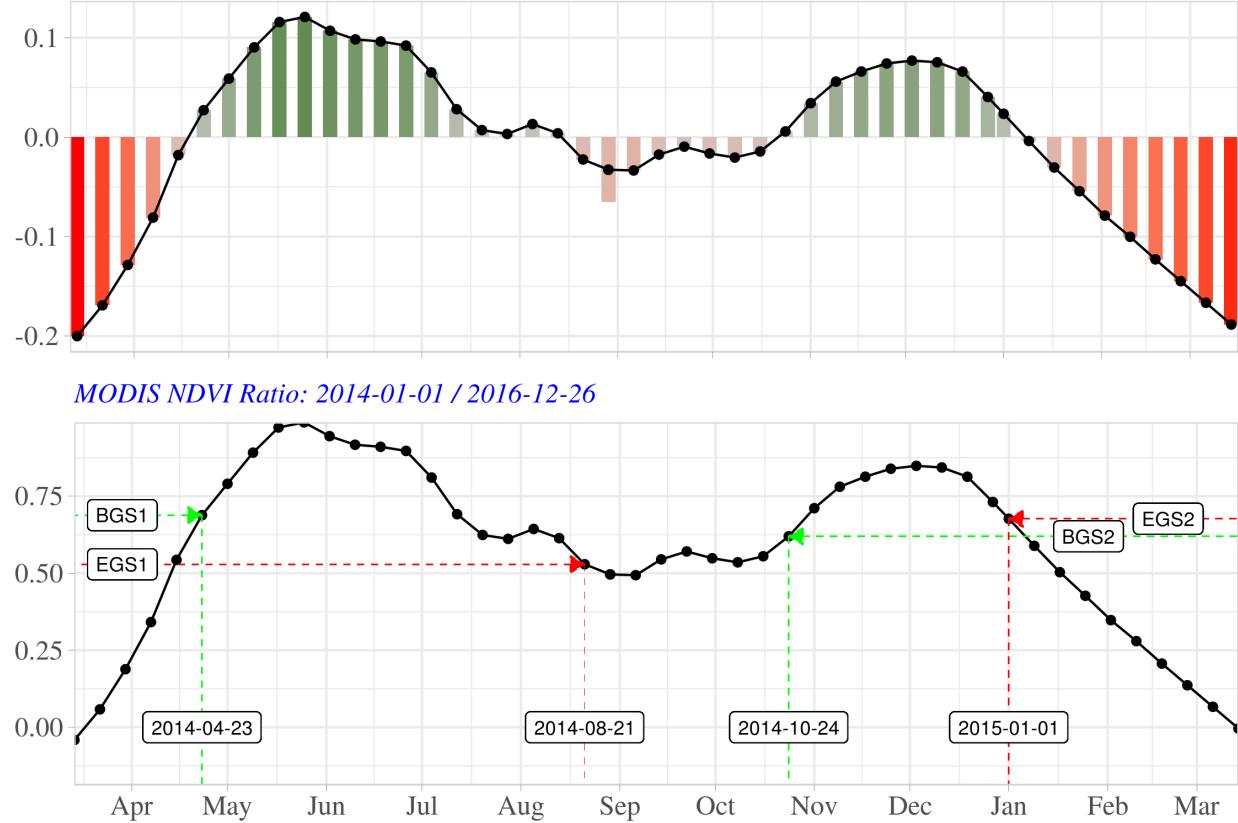


Figure 1: Vegetation seasonality in Kisumu County (Kenya).

Water presence

The water presence node estimates the likelihood of encountering water on the ground given the 3 years times series considered in this study. Flood is perhaps the most important characteristic of FBFS, and therefore water signature may be captured by the sensor in these areas. We therefore used the [NDFI](#) to assess such cases across all pixels in the study area. This was done using the procedure described in the main paper material.

Temporal variation in vegetation

The temporal variation in vegetation was estimated as pixel level vegetation anomalies using [NDVI](#) based on an approach of vegetation phenology (Figure 1). Our approach consisted in estimating the length of the growing season from which was then extracted an average interpolated surface derived from it. This is based on the assumption that the vegetation period is relatively extended under FBFS-like settings where plants take advantages of residual moisture from previous flooding.²⁵ In this case, these FBFS may be found in areas having unusually higher growing period relative to a general spatial trend.

EGS = End of growing season; BGS = Beginning of growing season; the indices (1 and 2) indicate the first and second seasons respectively.

The NDVI has been used in many ways to study vegetation phenology.^{26–28} We used the NDVI time series to first estimate the ratio (Figure 1) at which the vegetation density is changing throughout the year.^{27–29} The NDVI ratio was calculated using the Equation (2).^{27,29}

$$NDVI_{ratio} = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad (2)$$

Where $NDVI_{ratio}$ is the NDVI ratio of the pixel throughout the time series ranging from 0 to 1, $NDVI$ is the NDVI of that pixel at a given time step in the times series, $NDVI_{max}$ and $NDVI_{min}$ are respectively the maximal and the minimal NDVI values scored by the pixel over the time series.

The NDVI ratio (Equation (2)) is consistent since its value is normalised relative to itself. This ratio was derived considering the 3 years period considered in the study whereby the minimum and the maximal values were derived according to the phenological phases (Figure 1). An NDVI ratio at a given development stage informs on the proportion of vegetation relative to the total attainable biomass regardless of the land cover type.²⁹ While this method can help to avoid the use of locally adopted thresholds²⁹ since a generic threshold above which the onset and cession of the rainy season can be consistently estimated, deciding which ratio value to consider as that common threshold is another question. To address this concern, we extracted the seasonal component of the time series from which individual seasons were separated. Based on the assumption that the NDVI changes rapidly in the near onset and secession of the rainy season, we then scanned each season forwards and backwards questing for possible jumps (i.e. ± 3 standard deviation) in the values of the NDVI ratio. When scanning forwards, then a jump corresponds to the onset whereas the cessation is estimated scanning backwards. After locating the dates corresponding to these onset and cessation for each cell, their difference was calculated to estimate the length of the growing season for each season. These raster layers were aggregated to 5 km² interpolation surface using the `aggregate` function from the `raster` package¹³ to fit a thin plate spline regression model using the `Tps` function from the `fields` package.³⁰ The model was then used to interpolate each season using the `interpolate` from the `raster` package relative to the average interpolation surface.^{13,30} The anomalies were estimated as the difference between the length of the growing season and its interpolation. Finally, these anomalies were sum-up to provide a single estimate which was then discretised to provides nodes states for the Bayesian network as described in single layer procedure described in the main paper.

Flood at the beginning of the rainy season

The beginning of the rainy season node was defined widely from the starting of the season to nearly the vegetation peak to account for early and late flooding. The node flood at the beginning of the rainy season, then, estimates the probability of getting flood during that critical period. Based on this assumption, we extracted, from the NDFI times series, all

layers corresponding to the first 2 months from which the node states were derived using the multilayers procedure described in the main paper material.

III. Results and discussion

III.1. Comparative profiles of the NDSIs across different surface features

III.1.1. Comparative behavior of the NDSIs across biomes

The temporal profile of different NDSIs across the main ecological systems is provided in Figure 2. The NDSIs seem to behave similarly across the 3 main agricultural systems (i.e. Rainfed fields, Sugarcane fields, FBFS fields) with more pronounced flood peaks under FBFS fields. These peaks are also observed under flood-prone forests where floods seem to be seasonal. Rainfed fields seems to have more communalities (i.e. NDSIs' scores and temporal profile) with sugarcanes fields as do FBFS fields and flood prone forest although it is hard to distinguish between these 4 biomes.

While one may expect [NDFI](#) values above 0.5 at least under FBFS fields, the 3 water-related NDSIs (including the Gao [NDWI](#), Mcfeeters [NDWI](#)), [NDFI](#) scored negative values across the 3 different agricultural systems making thresholds specification difficult. The profile of vegetation ([NDVI](#), [NDII6](#), [NDII7](#)) under rainfed and FBFS fields seems to present a more pronounced seasonality contrary to sugarcane fields owing to their longer growing season resulting in moisture stability (Gao [NDWI](#)). This vegetation seasonality along with the extended length of the growing season becomes clearer under flood prone forest. Interestingly, flood prone forest seems to be different from FBFS fields with regards to vegetation density and open water stability. While both FBFS fields and flood prone forest scored positive [NDVI](#) values reaching up to 0.75, the [NDVI](#) values under flood prone forest rarely drop below 0.5 contrary to FBFS where these values can drop to 0.25. Under lakes systems (shallow and deep lakes), the signatures of the different NDSIs is rather confusing except in the case of [NDVI](#) which seems to be unique under shallow lake. While vegetation is high in the littoral zone, there seems to be a permanent mixture of vegetation and water in the deep water suggesting possible Eutrophication of Lake Victoria. Forest biomes are possibly distinguishable looking at the relative scores of the different NDSIs and their temporal profiles. Flood-prone and moderately dense forests are more seasonal in vegetation than dense forests exhibiting constant moisture over time. In general, sudden drop in vegetation seems to be associated with sudden increase in water signature under both dense and moderately dense forests while no such association is observed under flood-prone forest and towns suggesting the presence of open waters that are often masked by tree canopies. While vegetation ([NDVI](#), [NDII6](#), [NDII7](#)) decrease from dense forest to town through moderately dense forest as expected, the opposite trend is observed with open water. Nonetheless, the soil moisture (Gao [NDWI](#)) across the 3 biomes seems to be constant. Overall, most of the NDSIs behave similarly and open water stability or the length of the growing season can be misleading in delineating

Kisumu County, Kenya

MODIS : 2014-01-01 / 2016-12-26

— NDFI — NDII7 — NDWI (Gao, 1996)
 — NDII6 — NDVI — NDWI (McFeeters, 1996)

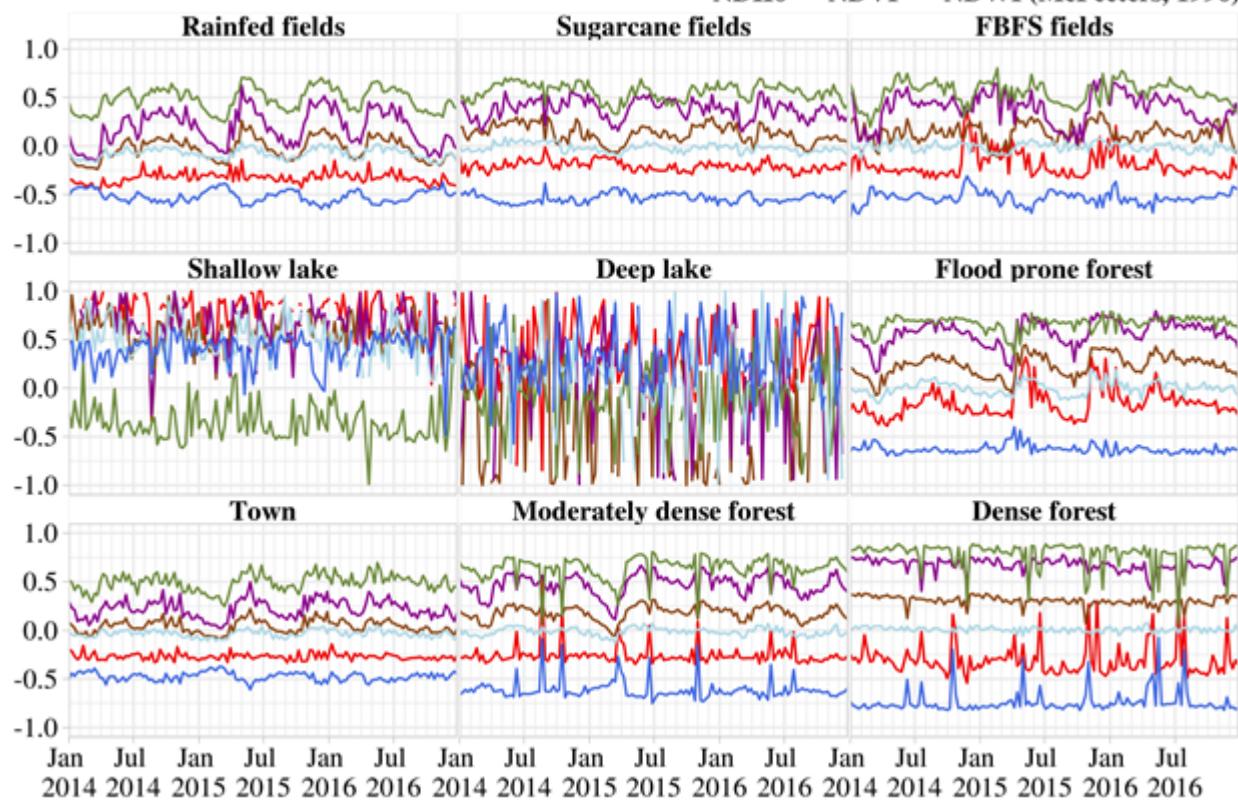


Figure 2: Comparative analysis and behavior of the NDSIs across different surfaces in Kisumu County (Kenya).

Kisumu County, Kenya

MODIS : 2014-01-01 / 2016-12-26

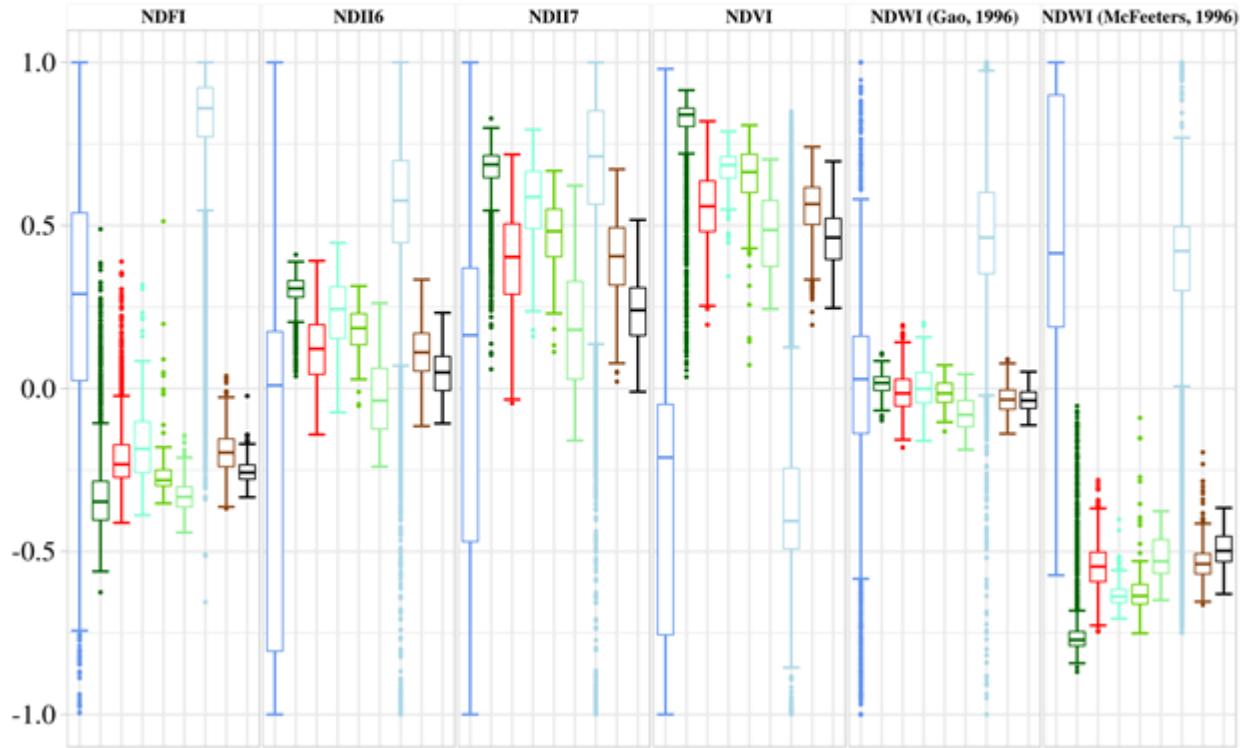
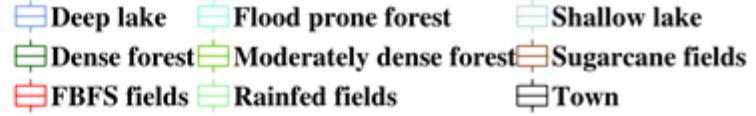


Figure 3: Comparative distribution of the NDSIs values across different surfaces in Kisumu County (Kenya).

FBFS fields in the area. Soil moisture (Gao NDWI) seems to be constant over time with little difference across biomes suggesting the presence of good soils for FBFS in the areas.

III.1.2. Comparative distribution of the NDSIs values across different surfaces

Looking at the distribution of the different NDSIs across the different biomes (Figure 3), it is clearly difficult to derive sharp threshold values for delineating most of the biomes, particularly the FBFS fields. The response of open water with regards to the different NDSIs seems to be rather odd, possibly due to the quality of the data since poor-quality data were associated with water bodies. Based the NDFI, the NDVI and the Mcfeeters' MDWI, FBFS fields can only be differentiated from shallow lake but the amount of noise is yet to be substantial considering the ranges overlapping of outliers. Considering the NDII6, the NDII7 and the Gao's NDWI, little can be done to distinguish FBFS fields from the other biomes due to the general lack statistical difference.

The best statistical difference based on the NDFI seems to be between the signature of the shallow lake and the rainfed fields. However, the majority of these rainfed fields seems to

experience flooding owing to their overlaps with both FBFS fields and deep lack. This is also observed with the Mcfeeters' MDWI. In a nutshell, there is redundancy in the scores of the NDSIs across the main biomes encountered in the study areas, hence they might be misleading for specifying thresholds. Consequently, there is a general lack of statistical evidence for deriving threshold values for delineating FBFS fields from the other type of ecological systems.

III.2. Phenological overview of the landscape texture

The average trend (represented by the interquartile range or the green medians line) depicted the phenological nature of the vegetation in the study areas with 2 annual phenological phases. The first phase seems to start from the beginning of April and ends during the period end of August – beginning of October whereas the second starts during the period beginning of September to the middle of October and ends in the middle of March. This main trend seems to isolate the croplands in relation with rainfall seasonality in the area (Figure 4).

The Outliers seems also to describe further phenological differences in vegetation across different type of biomes in the areas. Further examination of these outliers (not shown here) as separate datasets resulted in further temporal patterns in the NDVI values which can also be seen in Figure 4 (black and red lines). The first pattern (upmost outlier in black) seems to describe relatively high vegetation spots with NDVI values ranging from 0.5 to 1 during the period February – May. These could be due to data irregularities (sensor oversaturation during data recording, poorly corrected data, etc.) or more intuitively a sudden increase in vegetation due to the short rains in areas with relatively dense vegetation such as forests. In areas with permanent vegetation, changes in the greenness at the onset of the rainy season can be more rapid compared to agricultural fields where a certain time is required for crops to emerge and regreen the landscape. This is plausible considering that these outliers disappear when considered within the context of the upmost trend (upper whisker and upmost black line). The second pattern (lower outliers, and black line), accounting for vegetation indices in-between negative 0.25 and 0.5 seems to be positively correlated with the main pattern contrary to the third pattern (lower outliers and red line) describing vegetation indices below negative 0.25. while it is hard to tell the land use type described by the second pattern of these outliers, the third one is likely to describe water and water bodies. Overall, the study area is heterogenous and composed of several distinctive biomes with specific phenology and vegetation density, with some of these correlated over time.

III.3. Spatio-temporal variability in water and vegetation

The analysis of the NDSIs across the main existing land use / land cover provide little guidance to specify the boundaries of FBFS fields. Despite the differences in water and vegetation density (Figure 4), it was interesting to look at the main type of landscape existing in the areas and the extent to which their ecology can be appreciated in the context of seasonality. The PCA suggests that the first 3 components best describe the essential of the variability in vegetation supporting the possibility of data redundancy stated earlier

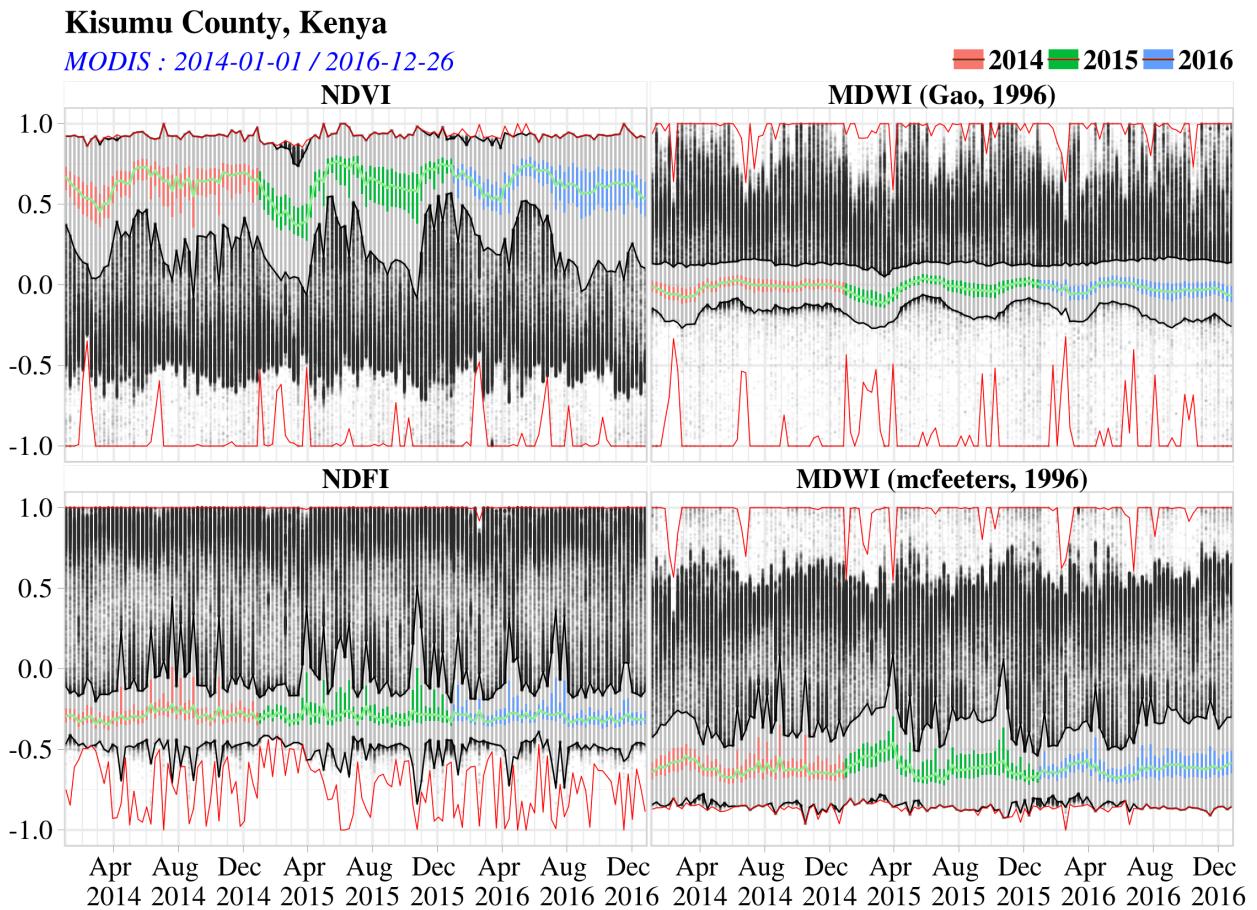
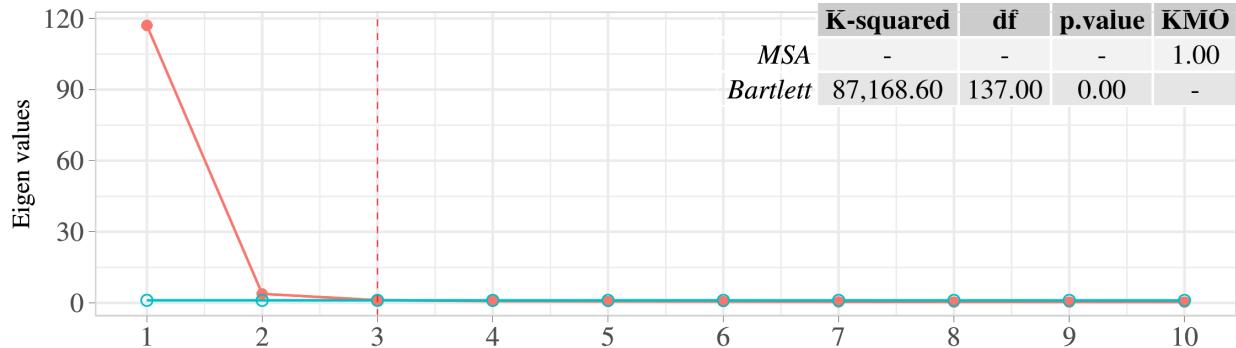


Figure 4: Boxplot of the temporal variability in water and vegetation in Kisumu county, Kenya.

Kisumu County, Kenya

MODIS NDVI : 2014-01-01 / 2016-12-26

Actual Data Simulated Data



MODIS NDFI : 2014-01-01 / 2016-12-26

Actual Data Simulated Data

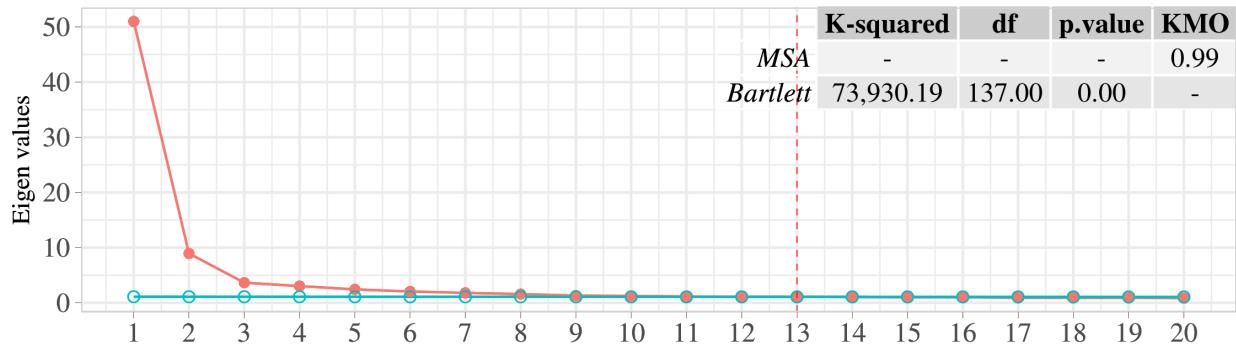


Figure 5: Scree plot of the first meaningful principal components along with the KMO and Bartlett's Test of sphericity

(section III.1.2.) whereas up to 13 were required for flood supporting the unpredictable nature of floods (Figure 5).

The spatio-temporal patterns of water and vegetation based on the 8 days MODIS composite in the study area presented in Figure 6 and Figure 8 respectively. The temporal variability describes the cross-pixels dominant variation over time attributed to the PCs depicted by the PCA. These are also presented as monthly aggregates to provide lumped view of the area. The Spatial patterns resulting from the scores of these PCs is also provided as maps. Based on these results,

Kisumu County, Kenya - 2014-01-01 / 2016-12-26

Temporal variability / Principal components (PC)

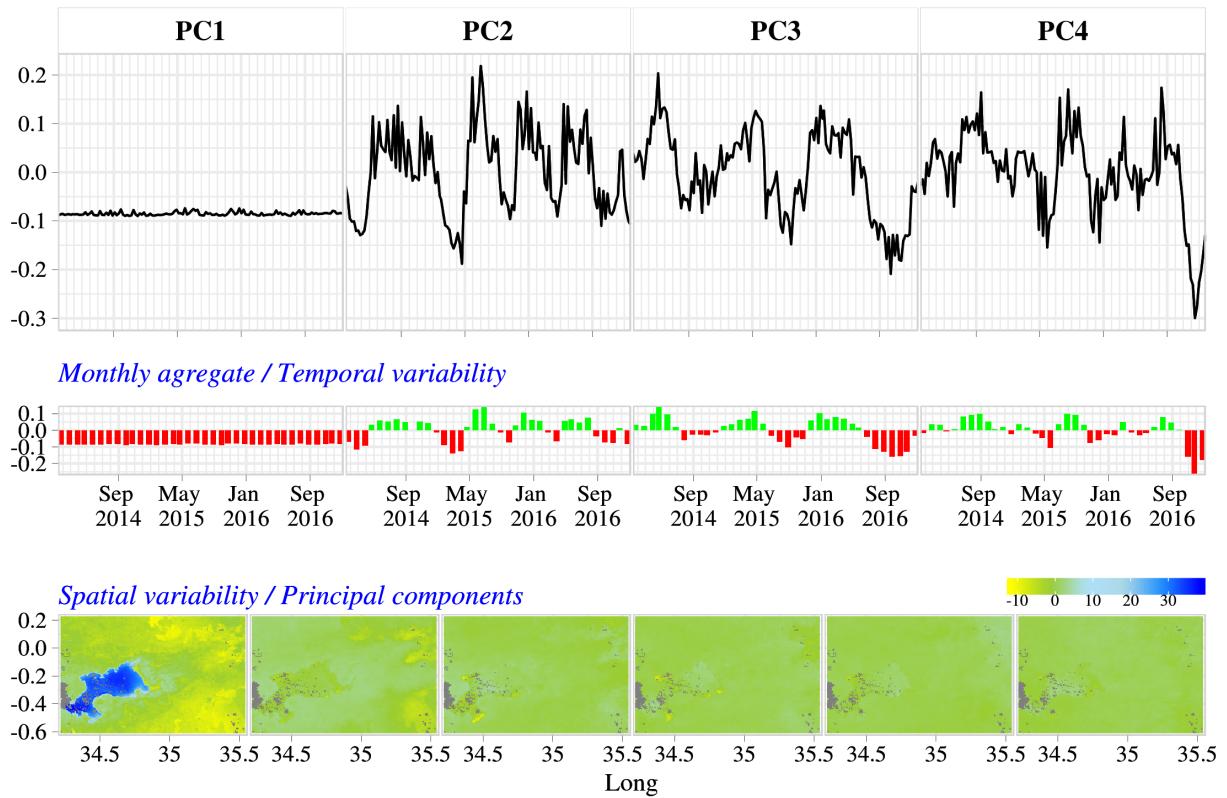


Figure 6: Dominants spatial and temporal characteristics of flood in Kisumu county, Kenya.

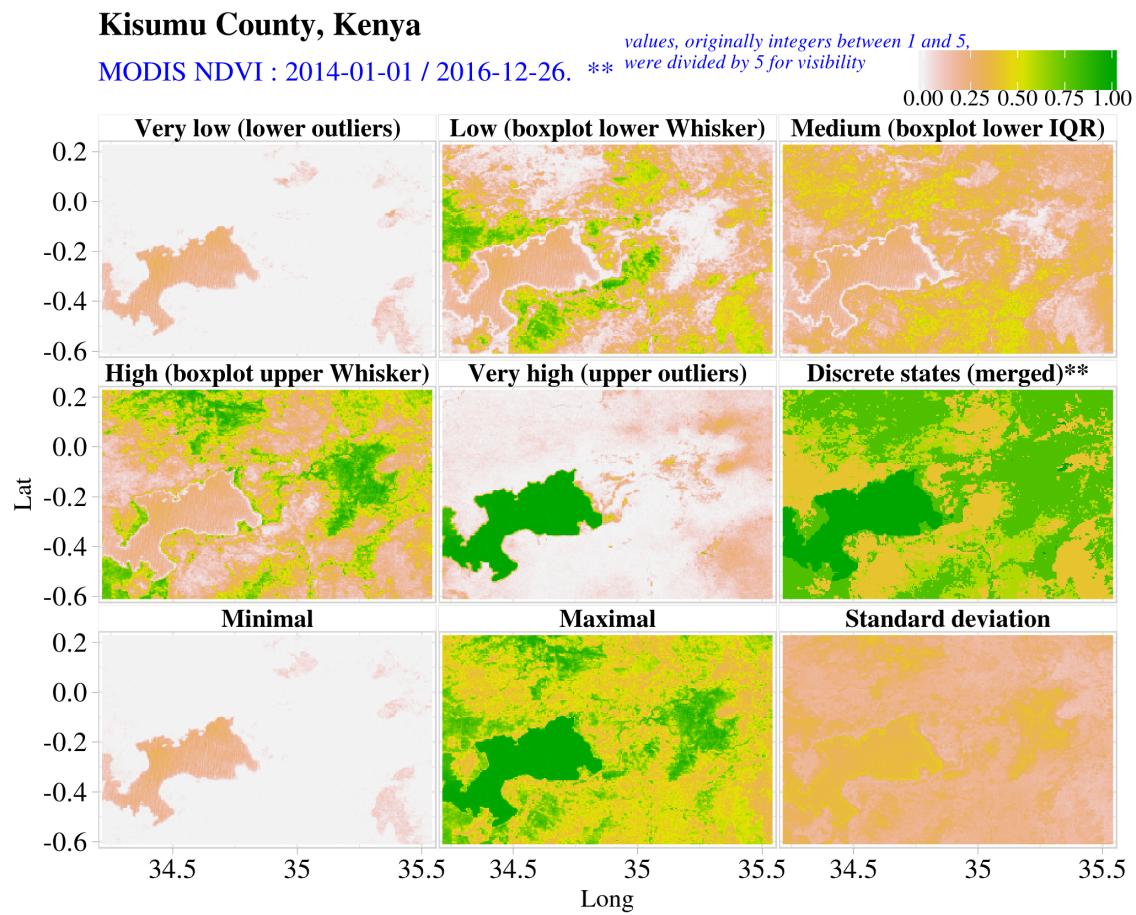


Figure 7: Probability of water by ranges in Kisumu county, Kenya.

Kisumu County, Kenya - 2014-01-01 / 2016-12-26

Temporal variability / Principal components (PC)

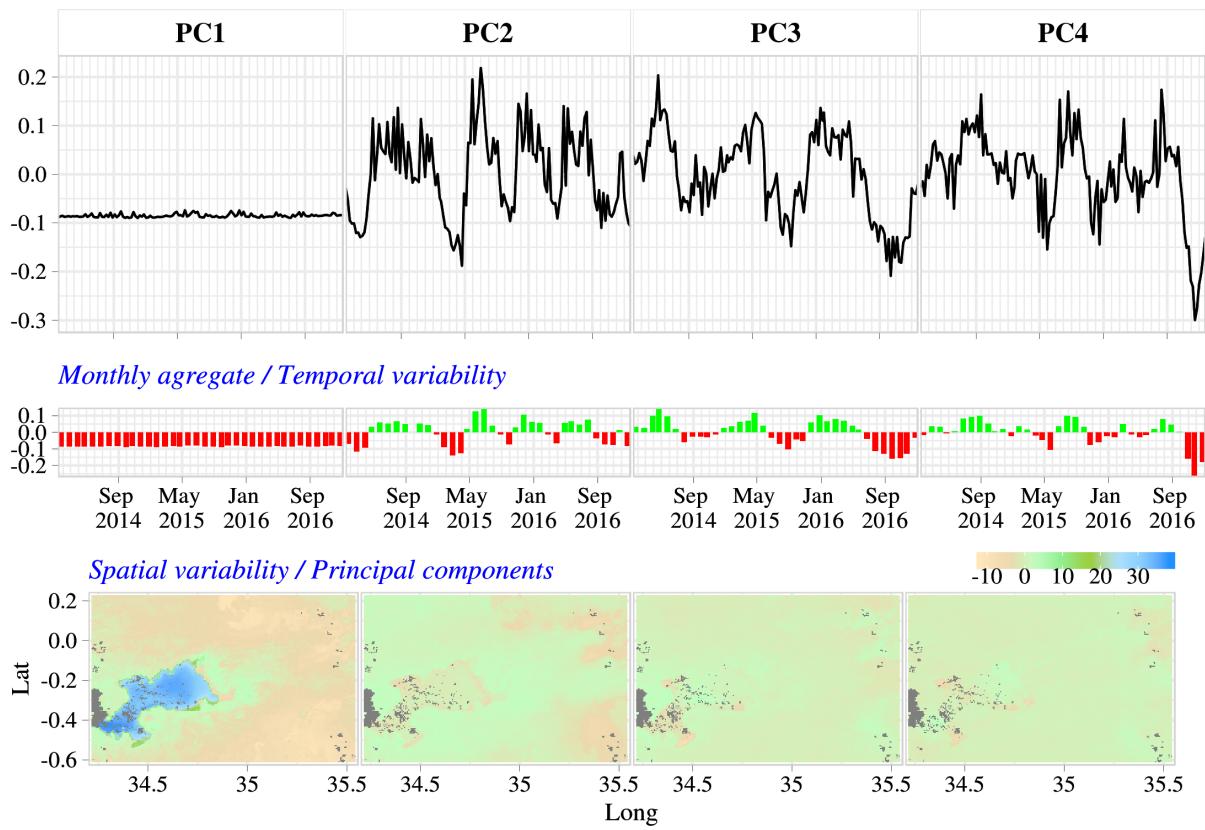


Figure 8: Dominants spatial and temporal characteristics of vegetation in Kisumu county, Kenya.

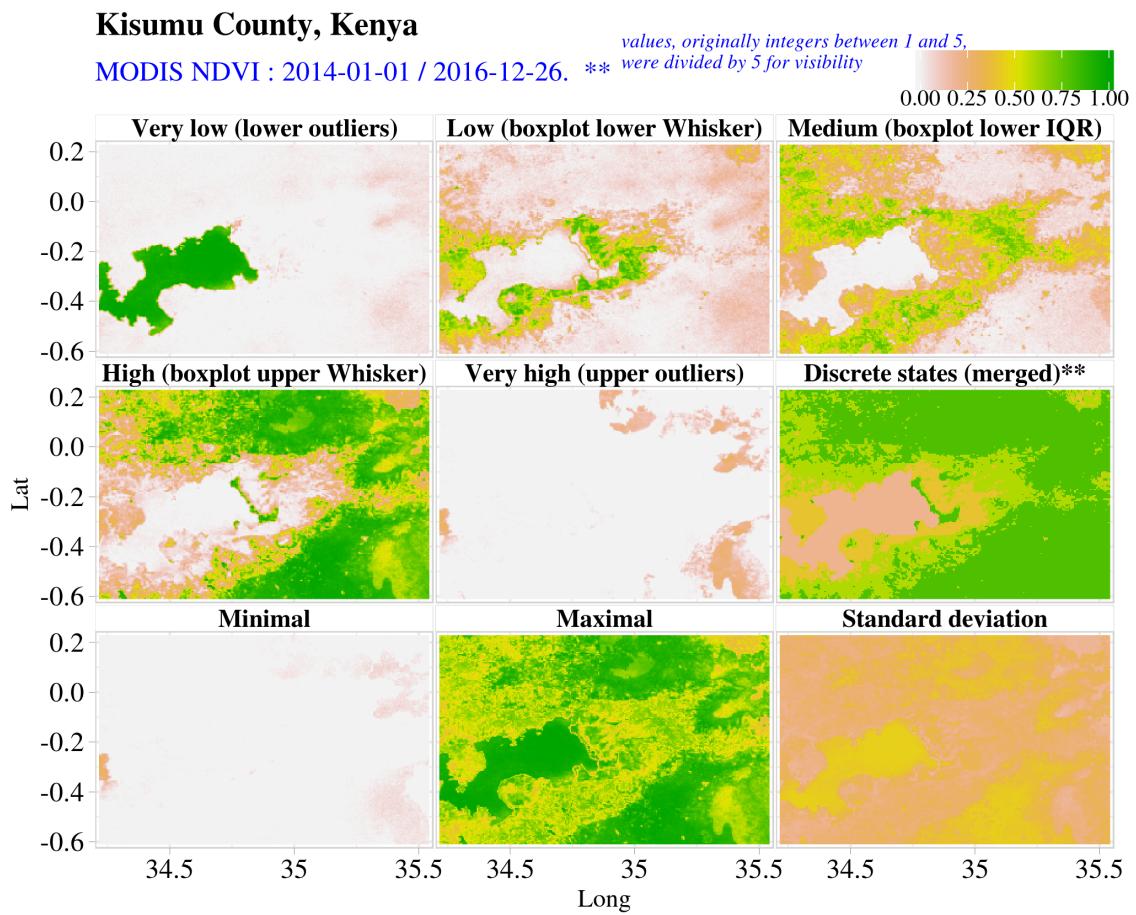


Figure 9: Probability of vegetation by ranges in Kisumu county, Kenya.

References

1. Bivand, R. S., Pebesma, E. & Gómez-Rubio, V. *Applied Spatial Data Analysis with R: Second Edition*. 405 (Springer New York, 2013).
2. Pebesma, E. & Bivand, R. S. S classes and methods for spatial data: the sp package. (2005).
3. Rouse, W., Haas, R. H. & Deering, D. W. Monitoring vegetation systems in the Great Plains with ERTS, NASA SP-351. in *Third earth resources technology satellite-1 symposium* 309–317 (NASA, 1973).
4. Tucker, C. J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment* **08**, 127–150 (1979).
5. Boschetti, M., Nutini, F., Manfron, G., Brivio, P. A. & Nelson, A. Comparative analysis of normalised difference spectral indices derived from MODIS for detecting surface water in flooded rice cropping systems. *PLoS ONE* **9**, (2014).
6. Gao, B. C. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* **58**, 257–266 (1996).
7. Ji, L., Zhang, L. & Wylie, B. Analysis of Dynamic Thresholds for the Normalized Difference Water Index. *75*, 1307–1317 (2009).
8. McFeeters, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* **17**, 1425–1432 (1996).
9. Mardia, K. V., Kent, J. T. & Bibby, J. M. *Multivariate Analysis*. 521 (Academic Press., 1979).
10. Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**, 559–57 (1901).
11. Leutner, B., Horning, N. & Schwalb-Willmann, J. RStoolbox: Tools for Remote Sensing Data Analysis. (2019).
12. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2018).
13. Hijmans, R. J. raster: Geographic Data Analysis and Modeling. (2019).
14. Hunt, E. R. & Rock, B. N. Detection of changes in leaf water content using Near- and Middle-Infrared reflectances. *Remote Sensing of Environment* **30**, 43–54 (1989).
15. Roy, D. *et al.* The MODIS Land product quality assessment approach. *Remote Sensing of Environment* **83**, 62–76 (2002).
16. Arge, L. *et al.* Efficient flow computation on massive grid terrain datasets. *GeoInformatica* **7**, 283–313 (2003).
17. Jenson, S. K. & Domingue, J. O. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing* **54**, 1593–1600 (1988).

18. Tarboton, D. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research* **33**, 309–319 (1997).
19. O'Callaghan, J. F. & Mark, D. M. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, & Image Processing* **23**, 323–344 (1984).
20. Tarboton, D., Bras, R. L. & Rodriguez-Iturbe, I. On the extraction of channel networks from digital elevation data. *Hydrological Processes* **5**, 81–100 (1991).
21. Tesfa, T. K. *et al.* Extraction of hydrological proximity measures from DEMs using parallel processing. *Environmental Modelling and Software* **26**, 1696–1709 (2011).
22. Yang, X., Chapman, G. A., Young, G. A. & Gray, J. M. Using Compound Topographic Index to delineate soil landscape facets from Digital Elevation Models for comprehensive coastal assessment. in *MODSIM 2005 international congress on modelling and simulation. Modelling and simulation society of australia and new zealand* (2006).
23. Tarboton, D. Terrain analysis using digital elevation models in hydrology. in *23rd esri international users conference* (2003).
24. Wallis, C., Watson, D., Tarboton, D. & Wallace, R. Parallel Flow-Direction and Contributing Area Calculation for Hydrology Analysis in Digital Elevation Models. in *The 2009 international conference on parallel and distributed processing techniques and applications* (2009).
25. Steenbergen, F. van, Lawrence, P., Mehari, A., Salman, M. & Faurès, J.-M. *Guidelines on spate irrigation*. 249 (FAO, 2010).
26. De Leeuw, J. *et al.* Benefits of Riverine Water Discharge into the Lorian Swamp, Kenya. *Water* **4**, 1009–1024 (2012).
27. Yu, H., Luedeling, E. & Xu, J. Winter and spring warming result in delayed spring phenology on the Tibetan Plateau. *Pnas* **107**, 22151–22156 (2010).
28. Yu, H., Xu, J., Okuto, E. & Luedeling, E. Seasonal Response of Grasslands to Climate Change on the Tibetan Plateau. *PLoS ONE* **7**, (2012).
29. White, M. A., Thornton, P. E. & Running, S. W. A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles* **11**, 217–234 (1997).
30. Nychka, D., Furrer, R., Paige, J. & Sain, S. fields: Tools for Spatial Data. (2018).