

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

Issoufou Liman^{1,2,}*

Cory Whitney^{1,3}

James Kungu²

Eike Luedeling^{1,3,4}

October 07, 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 was used to describe and interlink various features of FBFS using expert knowledge. 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 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 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. into mapping overlays. A Bayesian network ([BNs](#)) 3 years (2014-2016) of high-level [MODIS](#) were later used to feed the spatial data nodes of the [BNs](#) representing node states in [BNs](#). When such metrics are available, [BNs](#)

Contents

| | |
|--|----------|
| I. Introduction | 2 |
| II. Material and methods | 6 |
| II.1. Conceptual framework | 6 |
| II.2. Study area | 6 |
| II.3. Sensor Characteristics | 6 |

| | |
|--|-----------|
| II.3.1. SRTM | 6 |
| II.3.2. MODIS | 7 |
| II.4. Data acquisition and pre-processing | 7 |
| II.4.1. The BNs | 7 |
| II.4.2. The shapefiles data | 9 |
| II.4.3. MODIS Tera | 10 |
| II.5. Making sense of data in the context | 10 |
| II.6. Derivation of spatial data nodes | 10 |
| II.6.1. General Procedure | 10 |
| Time series data | 10 |
| Single layer data | 13 |
| II.6.2. Specific procedure | 13 |
| III. Results and discussion | 14 |
| III.1. Prior spatial distribution of the spatial data node states | 14 |
| III.2. Posterior spatial distribution of the spatial data node states | 17 |
| III.3. Case study: Application to Tigray region of Ethiopia | 18 |
| III.4. Potential of FBFS in Kisumu and Tigray | 18 |
| References | 21 |
| 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: <a href="mailto:issoufouliman<issoufoul@gmail.com>">Issoufou Liman <issoufoul@gmail.com> | |

I. Introduction

Flood-based farming systems (FBFS) are rainfed farming systems occurring in drylands areas and receiving extra irrigation from various type of non-harmful floods that share the properties of being unpredictable, of short duration, of low annual frequency and occurring in relatively lowlands areas with gentle local topography where various social and institutional arrangements govern the water access and sharing.¹⁻³ By making flood water available for

use in agriculture, these farming systems contribute to food security along with many other benefits for drylands' society.^{2,4} They are mostly extensive subsistence-based but support millions of farmers especially in East Africa where they also provide income and various other intangible services.¹ Many countries across Africa and Asia endorsed their development through a common framework for research, policy and action (i.e. Flood-based Livelihood Network (FBLN) foundation) mandated for knowledge sharing, filling critical knowledge gaps and improving their productivity of these farming systems.^{1,5}

Despite their promises and importance, FBFS have been the topic of surprisingly few studies² focusing on aspects related to hydrology and sedimentation, design and maintenance, or sociology attuned to specific contexts.^{2,6} Consequently, there is a general lack of information to answer relevant questions for FBLN countries. For example, little effort has been made to provide reliable estimates of their coverage.² Despite the existing country-specific area estimates, it is often unclear how these were derived resulting in large uncertainties across these estimates.² Most FBLN countries' reports expressed the need for mapping FBFS, yet only few attempts^{7–9} are available towards this end in our knowledge. While most aspects related to FBFS are unpredictable,^{1,2} these approaches seem to rely on the assumption that agronomic flooding can be clearly detected on a satellite image in the near time of flooding event and sometimes imply that the areas to be mapped and the timing of the flooding is already known and matches a remote sensing scene in space and time. Furthermore, some of these approaches rely on the derivation of some kind of thresholds based on which FBFS can be discriminated. Threshold-based methods may be impractical considering the complexity of FBFS, the communalities they share with both forests, open water, rainfed and irrigated systems,^{1,2} and particularly when discriminating between different type of FBFS.¹⁰ Moreover, some of these methods rely on visual interpretations which are hard to upscale and may require a keen knowledge of the area. While it is challenging to map these FBFS, accurate information is needed to monitor them, assess their contribution to food security, biodiversity conservation and other intangible benefits at various scales, urging the need for standard methodologies for mapping them. We argue such a surveillance can be best achieved using Geoinformation.

¹¹ covered a range of capabilities and applications of remote sensing in development research, including the use of spectral vegetation indices, and digital elevation models in ecological studies which have potentials in mapping FBFS. Even though satellite remote sensing provides a framework for comprehensive mapping of these farming systems,^{11,12} traditional methods may fail to discriminate them due to their complexity and diversity across different settings.¹⁰ The complexity arises from their similarities with both rainfed and irrigated systems^{1,2} along with a number of their biophysical attributes that could exhibit different signatures and can, therefore, only provide partial information for discriminating them. In flood-based agriculture, one may expect, for instance, an open water followed by rapid increase in vegetation or a combination of thereof.¹⁰ The sequence open water-vegetation, though, can be misleading because the flood event can take place without leaving detectable trace of inundation not only because the sensor can miss to capture the flood event but also because many FBFS have deep soils capable of storing large amount of water. The flood event may also occur after the planting date in which case it can be confused with other flood hazards. The diversity of flood-based agriculture may pose another challenge for

mapping FBFS because locally adopted thresholds may be required¹³ due to the diversity of conditions under which they are practiced.^{1,2} In addition, the areas under the practice may be unstable over time because of changes in the water courses, water flow, etc..² There is clearly a need for generic approaches for mapping FBFS which can be applied to several contexts given their diversity and the diversity of crops grown under them.¹

Mapping such complex systems, where events can be causally related in a context of data uncertainty may require multivariate approaches that are able to leverage on both available information and expertise.^{14–17} Considering the data uncertainty for remote sensing of FBFS, Such an approach, conversely, needs to be able to combine knowledge from domain experts with learning from all available data, be suitable for incomplete information while accounting for the sources of uncertainties.^{18–20} These expectations can be achieved using BNs.^{20,21} The technicalities along with the many real-world applications of BNs are beyond the scope of this paper and are extensively documented elsewhere.²²

The objective of this paper is to provide an affordable mapping routine that could be used to comprehensively estimate the coverage of FBFS. However, the concept is reproducible in other contexts where spatial data are available. Based on the definition and characteristics of FBFS,^{1,2,23} examining a number of biophysical surface states related to their characteristic hydrology, vegetation, and topography, preferably in forms of time series where possible, are clearly relevant for mapping them. With this assumption, we used open source satellite remote sensing data and software to capture various FBFS attributes following a conceptual model developed by experts using various normalized difference spectral indices (NDSI) and topographic structures^{11,24–34} specific to FBFS to ultimately derive their probability of occurrence for each pixel across Kisumu County in Kenya. We argue that threshold-based methods may be unsuitable to mapping such complex systems and demonstrate the versatility of probabilistic methods. Despite the ranges of data uncertainty under such complex systems, we elaborate on the existence of some sort of spatio-temporal pattern in the general landscape that can be described using causality and from which fuzzy metrics can be derived to describe distinctives sub-entities. We described how this can be achieved and by deriving the required inputs for BNs from raw spatial data. The approach was tested across the Tigray region in Ethiopia (not presented here); an area with different settings; and seems to perform incredibly well. The core data used in the study include three years (2014–2016) of high-level Moderate Resolution Imaging Spectroradiometer (MODIS) version VI products acquired in the forms of NDSI, the Shuttle Radar Topographic Mission (STRM) void-filled Digital Elevation Model (DEM). Data analysis were conducted using the R programming language³⁵ and software packages are referred to as such unless otherwise specified.

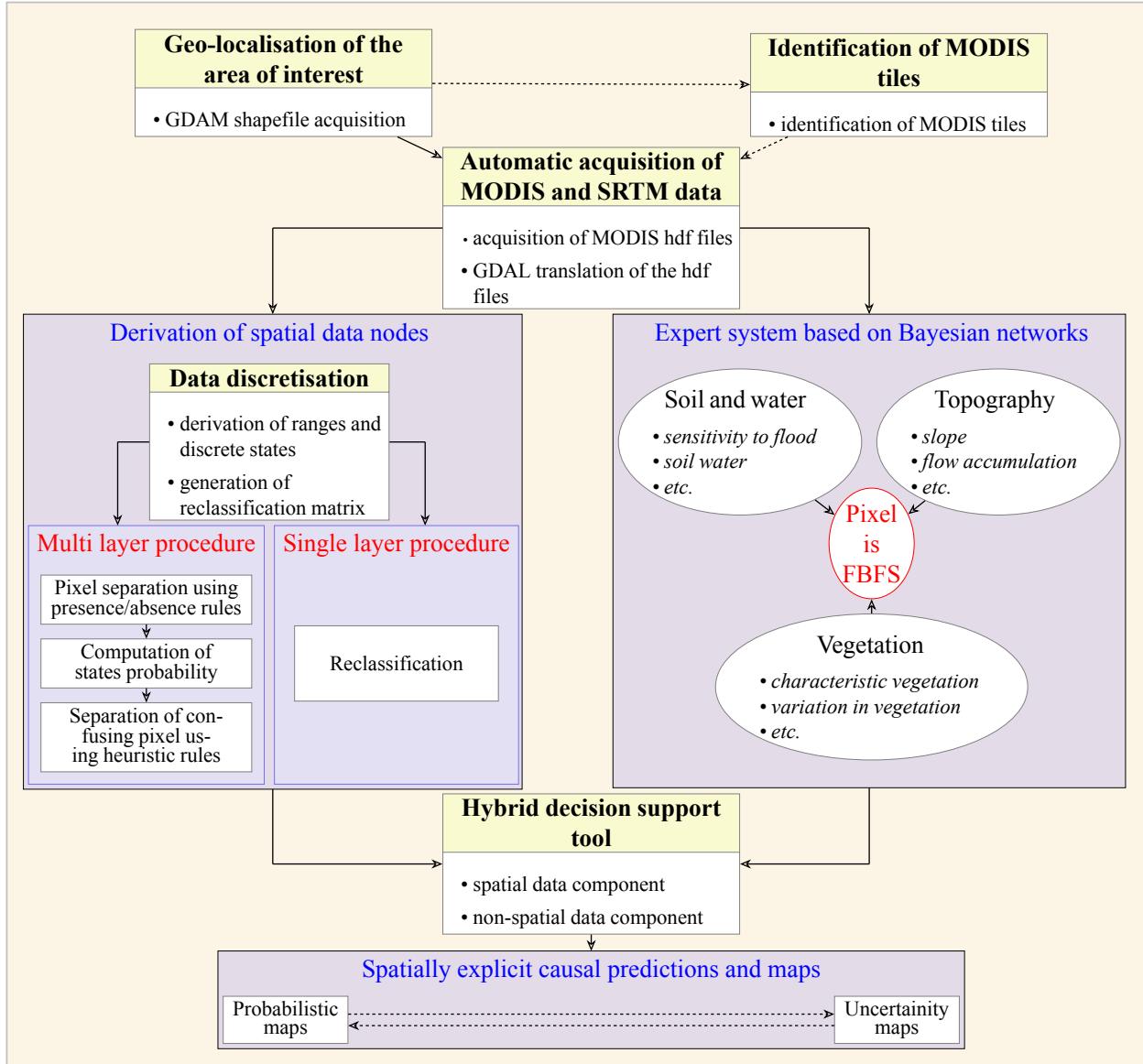


Figure 1: Conceptual framework for mapping FBFS.

II. Material and methods

II.1. Conceptual framework

II.2. Study area

The study took place in two regions of Kenya and Ethiopia, the Kisumu county and the Tigray region respectively. Both areas have long history in terms of the practice of flood-based agriculture and have different ways of managing agronomic flooding. However, this paper specifically focuses on the case of Kisumu to provide the backbones of the methodology. Kisumu is one of the largest of Kenya's counties, located in the western Kenya.³⁶ It is bordered by the Homa Bay County to the South, Nandi County to the North East, Kericho County to the East, Vihiga County to the North West and Siaya County to the West. Kisumu County extends between 33° 20'E and 35° 20'E and 0° 20'S and 0° 50'S covering an area of 2576.5 km² of which 567 km² (4.5%) is covered by water.³⁶ Common topographic features are the Winam Gulf of Lake Victoria and the Rift valley. The central part of the county is relatively flat lands surrounded by higher altitudinal ridges reaching up to 1835 m. The relief of the county can be described by 3 main topographical features, the Kano lowlands plains, the Maseno midlands, and the highlands of Nyabondo Plateau making most of the central part prone to flooding, particularly during periods of heavy rains.³⁶ Kisumu county is endowed in surface water with three main rivers flowing into Lake Victoria.³⁶ The floodplains are good for agriculture due to their relatively rich soils resulting from recurrent alluvial deposits and potential for irrigation.

II.3. Sensor Characteristics

II.3.1. SRTM

SRTM is an collaborative project between the US National Aeronautics and Space Administration (NASA) and the National Imagery and Mapping Agency (NIMA), the National Geospatial-Intelligence Agency, and the German and Italian Space Agencies providing the most complete highest resolution and public domain DEM between 56° S and 60° N.³⁷⁻³⁹ It uses a dual radar antennas interferometry to produce high resolution topographic data and carries over 300 high-density tapes storing these data. several spatial resolutions (e.g. 30m, 90m, 1000m) are available.^{38,40} A list of available SRTM data and their sources along with various other information can be found in Farr et al.³⁸ and the accuracy of the SRTM data is extensively discussed in Falorni et al.⁴¹. In this study, we used the 30 m spatial resolution DEM on which further processing were applied to ultimately compute the spatial data nodes used in the BNs.

II.3.2. MODIS

MODIS is a key remote sensing sensor on both Earth Observing System (EOS)-AM1 and EOS-PM1 operating onboard Terra and Aqua satellites respectively.⁴² MODIS acquires data in the visible and infrared regions of the electromagnetic spectrum to provide 36 spectral bands with a viewing swath width of 2,330 km for studying various aspects of terrestrial ecosystems. MODIS scenes are geo-localised using a sinusoidal projection system and therefore identified using their specific horizontal (h) and vertical (v) tiles number. These data are provided in different spatial (250 m, 500 m, and 1 Km) and temporal (e.g. daily, 8 days, 16 days) resolutions. In this study, we used the MODIS product MOD09A1 (Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006⁴³ which provides an 8-day composite atmospherically corrected land surface reflectance across bands 1 to 7 along with other additional quality indicators layers assessing the quality of the acquisition at pixel level.

II.4. Data acquisition and pre-processing

II.4.1. The BNs

BNs are reasoning engine and knowledge engineering tools that can be used to model partially known events in forms of probability, incorporate causal uncertainty in the analysis,²² and therefore allow reasoning under uncertainty. On the one hand, they are causal probabilistic models, on the other hand, they are ways of decomposing large joint probability distributions^{22,44} and specifying complex heuristics that go beyond a simple regression model.¹⁷ In Bayesian analysis, scientific reasoning starts with the current state of knowledge and subsequently assesses the contribution of further new information to that knowledge.⁴⁴ BNs become more exciting and have proven to give incredibly remarkable results when there is a need to incorporate many variables in a model and determine their effects on an outcome.

In this study, the ultimate outcome variable (`PixelIsFBFS`) determining whether a given pixel qualify for FBFS is described using causal relationships among several other variables specified in a BNs model (Figure 1). The conceptual model described pixels' potential for FBFS (`PixelIsFBFS`) to be jointly defined by the amount of flood water received (`FloodSomePts`), the topographic suitability, and the extent of vegetation variation due to flood (`VegVarFlood`) at that pixel. A pixel may record important run-on without experiencing flood because the water may run-off without a suitable topography (`suiteTopo`). This is, vegetation variation in space and time can be due to either flood or non-flood water but each pixel can be profiled in terms of its amount of flood water.

The topographic suitability of a pixel (`suiteTopo`) depends largely on its slope (`Slope`) and the number of pixels draining into it (`FlowAcc`). While the accumulation of flow (`FlowAcc`) is important to the amount of flood reaching the pixels, it is rather the slope (`Slope`) and the gravity that determine whether the water will be kept or transmitted by that pixel.

The extent to which vegetation variation is caused by flood water (`VegVarFlood`) can be a straightforward indicator for mapping flood-based agriculture but also difficult to measure.

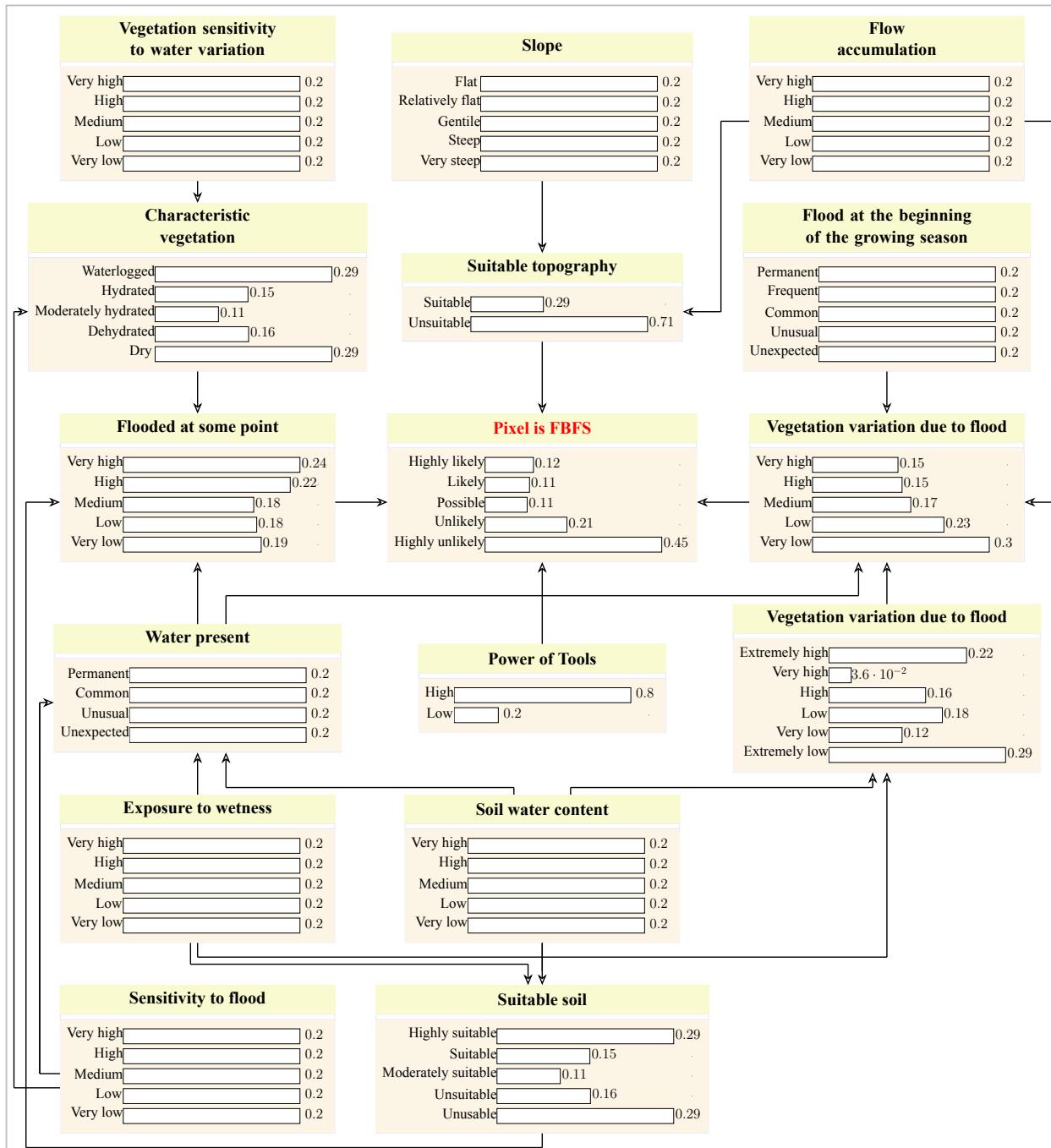


Figure 2: Bayesian Network describing the causal reasoning used for mapping FBFS in Kisumu, Kenya.

Table 1: List of normalized difference spectral indices (NDSI) used for mapping FBFS in Kisumu County, Kenya.

| | Bands | Spectral range | Full name |
|----------------|--------|---------------------------------|--|
| NDVI | B2, B1 | (NIR – VISRed)/(NIR + RED) | Normalized Difference Vegetation Index (Rouse, 1973) |
| NDFI | B1, B6 | (NIR - SWIR1) / (NIR + SWIR1) | Normalized Difference Flood Index (Boschetti, 2014) |
| NDII6 | B2, B6 | (NIR2 - SWIR1) / (NIR2 + SWIR1) | Normalized Difference Infrared Index - Band 6 (Hunt, 1989) |
| NDII7 | B2, B7 | (NIR2 - SWIR2) / (NIR2 + SWIR2) | Normalized Difference Infrared Index - Band 7 (Hunt, 1989) |
| GAO MDWI | B2, B5 | (NIR2 - NIR5) / (NIR2 + NIR5) | Normalized Difference Water Index (NDWI) (GAO, 1996) |
| McFeeters MDWI | B4, B2 | (VIS4 – NIR2) / (VIS4 + NIR2) | Normalized Difference Water Index (NDWI) (McFeeters, 1996) |

Vegetation variation caused by flood (`VegVarFlood`) was therefore assessed via proxy variables. For the temporal variation in vegetation indices (`TempoVarInVi`) to be attributed to flood to some extent (`VegVarFlood`) at a given pixel, one may expect a certain degree of flow accumulation (`FlowAcc`), and a certain likelihood of flood events near the beginning of the growing season (`FloodBgs`). Then, the temporal variation in vegetation (`TempoVarInVi`) of the pixel is assumed to be influenced by a number of factors such as its soil water content (`SoilwaterCont`), its exposure (`Expos2wet`) and sensitivity to flood (`Sens2Flood`), or even the quality of the data in relation with the type of algorithm used in processing these data (`PowerTools`).

For a given pixel to be flooded at some points (`FloodSomePts`), there must be water presence on the ground (`WaterPrznt`). This is moderated by the soils suitability (`SuitSoil`) resulting in specific type of vegetation (`CharactVeg`). Therefore, the extent to which a pixel is flooded (`FloodSomePts`) can be deduced from its typical vegetation (`CharactVeg`), the soil suitability (`SuitSoil`), and the presence of water (`WaterPrznt`). Soil suitability to FBFS is defined by its exposure to wetness (`Expos2wet`) and more importantly its water content (`SoilWaterCont`). Water presence (`WaterPrznt`) may depend on soil water content (`soilWaterCont`), the power of tools (`PowerTools`), and the sensitivity to flood (`Sens2Flood`). Characteristic FBFS' vegetations are assumed to be due to vegetation sensitivity to water variation (`Veg2waterVar`) and the sensitivity of the area to flood (`Sens2Flood`).

II.4.2. The shapefiles data

The shapefiles of the study areas were acquired from the [global administrative boundaries database](#)⁴⁵ as level 1 product, using the string country name as argument to `getData` function from the `raster` package,⁴⁶ from which we extracted the shapefile of the area of interest. These shapefiles were latter used to identify the `MODIS` H/V titles corresponding to the area. Other shapefiles used are the spatial polygons representing well known selected surface features (land use/ land cover) used to understand the behaviour of the spectral response in the area. These were either collected during field works using handheld GPS devices or digitized on Google Earth to be used to query the spectral response of pixels they match spatially. The surfaces features considered here include different depth water bodies, settlements, FBFS fields, rainfed agricultural fields, riparian forests, forests with varying degree of density, and sugarcane fields having relatively extended growing season.

II.4.3. MODIS Tera

We considered a period of three years (2014 - 2016) and queried all available **MODIS** Tera data overlapping the acquired shapefiles of the area of interest (section II.4.2.). We first determined the Horizontal and Vertical tiles corresponding to the spatial coverage of the shapefiles and used these to define the appropriate geo-locations of the study area relative to the **MODIS** Sinusoidal grid systems using the `getTiles` function from the **MODIS** package.⁴⁷ As result, the Kisumu County falls in-between the h21v09 and h21v08. Based on these tiles and the temporal window of interest, we mosaiced the images and calculated the NDSIs after filtering and downloading the data from MODIS global online database (i.e. LP DAAC, LAADS) using the machine-to-machine routines embedded in the `MODISstsp` function from the package of the same name.⁴⁸ Since the **MODIS** data are provided in hierarchical data format (hdf), they were re-projected to World Geodetic System 1984 (WGS 84) and subsequently translated into Geotiff format using `GDAL 2.2.3`.⁴⁹

II.5. Making sense of data in the context

Prior to processing the data acquired, preliminary analyses were conducted to understand the behavior of the spectral response with regards to water and vegetation in the study area. These analyses include the boxplot of the time series to understand the temporal variability and the principal component analysis to depict the spatio-temporal pattern. The results are presented in the supplementary materials.

II.6. Derivation of spatial data nodes

II.6.1. General Procedure

Time series data

The boxplots and the comparative analysis of the NDSIs shows different patterns but these can only be hardly separated making it difficult to threshold the different NDSIs across land use / land cover. An approach based on rigid thresholding was clearly not the best option to demarcate the nature of the spectral response, at least not without making complicated statistical assumptions. A simpler way would be to approach the problem using imprecise quantifiers where fuzzy logic can be used in forms of Likert-type scales⁵⁰ describing the scores of the spectral responses along a range; supporting the use of **BNs**. We therefore make use of the different ranges of the boxplot (i.e. lowest outliers – lower Whisker, lower Whisker - 1st quartile, 1st quartile – 3rd quartile, 3rd quartile - upper Whisker , upper Whisker – highest outliers) to separate the data and estimate the probability of each pixel falling within each

of these ranges at each time step in the time series.

Algorithm 1: make pixel states

Result: a spatially explicit quantitative representation of variable states using probability

Input : NDSI time series stack

Output: a raster stack where each layer represents one of the variable states

```

1 for each layer in the time series do
2   | compute Whisker ranges
3   | for each Whisker range do
4     |   | if cell value belongs to the Whisker range then
5       |   |   | recode the cell value to 1
6     |   | else
7       |   |   | recode the cell value to 0
8     |   | end
9     |   | compute the probability of getting 1
10    | end
11 end
```

Although rarely used for deriving variable states, boxplot has been used in all kind of statistics for the last 40 years. Beyond their incredible simplicity, boxplots use robust statistics to summarise data in terms of five-numbers while being particularly handy when comparing the distributions of these data across groups.⁵¹ To estimate the probability for a given range, we first recoded all values outside of that range to 0 whereas the values falling within the range were recorded to 1. This way, all the real-valued time series was reclassified into presence - absence time series data. Thereafter, the final probability of a pixel belonging to a range can be easily estimated using the ratio between the presence and absence of that pixel in the range throughout the time series (Equation (1)).

$$P(p_i \in range_j) = \frac{n_{p_i \in range_j}}{N} \quad (1)$$

Where $P(p_i \in range_j)$ is the probability of the pixel p_i belonging to the range $range_j$, $n_{p_i \in range_j}$ is the number of time the pixel p_i belonged to the range $range_j$, and N is the sample space (total size of the time series).

After estimating the probabilities for each boxplot range, these ranges were mapped into fuzzy linguistic quantifiers in forms of low, medium, high etc. (depending on the number of states identify by the boxplot algorithm) to provide vegetation states in a spatially explicit manner. For example, in situation where all the five boxplot ranges are present, lower outlier values were mapped to very low, the lower Whisker to low, the interquartile range to medium, the upper Whisker to high and the upper outliers to very high. These states, computed in separated raster layers, were then merged into one single layer to provide discrete states on

which the BNs can be easily operationalized.

Algorithm 2: compile pixel states

Result: a spatially explicit qualitative representation of variable states using common language

Input : NDSI time series stack

Output: a raster layer where pixels are represented using relative fuzzy linguistic quantifiers

```

1 do algorithm 1: make pixel states
2 for each pixel do
3   if the pixel has one single maximum for all layers then
4     assign to the pixel the position of the layer (corresponding to a Whisker range as
      in the stack generated at step 1 and mapping to a linguistic quantifier) holding
      that maximal (probability) value
5   else
6     assign NA to the pixel ;
7   end
8   for each pixel holding NA in the layer (generated after step 6) do
9     if the pixel has one single maximum for all neighbours in the 8 ways
       connectedness then
10    | fill it with that maximal value
11    else
12    | leave the pixel value as NA ;
13    end
14   get the position of the remaining pixel holding after step 12
15   extract all pixel values at the positions in step 12 from the stack generated at
      step 1
16   fit a classification regression tree model to the extracted data in step 15
17   predict the remaining pixels holding NA after step 12 using the model developed
      at step 16
18 end
19 end

```

The decision to assign a state value to a given pixel was made transparently using fully probabilistic heuristics in 3 sequential steps. The first step consisted in assigning a particular state to a given pixels if that state scored the highest probability and only if that highest probability is exclusive. Thus, all cases where 2 or more highest probability co-exist are considered as unknown and ignore until the second step. Such cases having equal probability states were commonly encountered at the edges where two consecutive states spatially meet. In the second step, these confusing pixels having 2 or more maximal probability of state occurrence are filled in using majority rule. Herein, since spatial feature are geographically auto-correlated in general, we applied a 3 by 3 moving window (eight-way connectedness) centred on each of these uncertain state value pixels and estimate their state value based on the values of the 8 neighbourhood pixels. Similarly, to the first step, a particular value among the values of the 8 neighbours is chosen when and only when it is the only value

that occurs the most. Whenever 2 or more states happen to be the most probable at the same time, the pixel is left out and handled at the third step. In the third step, the rest of these confusing pixels, where the first and second steps heuristics failed, were modelled using decision tree recursive partitioning classification algorithm. We first extracted the maximal values which correspond to the highest probability across all these remaining confusing pixels to fit a classification regression tree (CART) model and used that model to conditionally infer their state values. Therefore, after this third step all pixel state value were estimated, and the resulting raster can be used as spatial data node in the `BNs` model. The boxplot statistics were computed using our own defined function (see `get_boxplot_1d` in the technical material) based on `boxplot.stats` function from the `grDevices` package.³⁵ We used the `reclassify` and the `calc` functions both from the `raster` package⁴⁶ to recode the pixel values and to estimate the probability respectively. The second step confusing pixels were handled using `modal` function from within the `focal` function also from the `raster` package.⁴⁶ The CART model was fitted using the `cTree` function from the `party` package⁵² whereas the third step confusing pixels' inference was achieved by supplying an anonymous function wrapping the `predict` function from the `stats` package³⁵ as argument to `calc` function.⁴⁶

Single layer data

In the case of single layer data where time series does not make sense (e.g. slope and flow accumulation) or where nodes were derived from times series as single layer (e.g. exposure to wetness, vegetation sensitivity to water), a slightly different procedure compared to that adopted for time series data was adopted for `BNs` nodes making. We rather used the boxplot's five-numbers approach in a much straightforward way by exploiting the spatial variability across cell values since the temporal component of this variability no longer exists. We first extracted all pixel values and computed the five-numbers from which ranges corresponding to discrete states were computed. These ranges were then directly recoded into unique values mapping to fuzzy linguistic quantifiers. The process is somewhat a simple raster reclassification task except that the reclassification matrix was derived directly from the natural breaks of the data.

II.6.2. Specific procedure

Specific processing procedures were applied to derive each spatial data node considered in the `BNs`. These are described in detail in the supplementary material. We explained the meaning of each node, provided the rational of its usage and described how it is causally related to other nodes in the `BNs`. Some nodes (e.g. Soil water content) were indirectly estimated as composite variable involving several NDSIs. Other (Exposure to wetness) were simply computed using single metric time series. In some other cases, the computation involved complex algorithm applied either to time series data (Temporal variation in vegetation) or to single layer (Flow accumulation) depending on the nature of the data.

Table 2: Spatial coverage of different states of vegetation and flood in Kisumu County, Kenya.

| | Very low | Low | Moderately Low | Moderately High | High | Very high |
|------------|-----------------|-----------------|-----------------|-----------------|----------------|-----------------|
| Vegetation | 544.22 (20.42%) | 114.51 (4.3%) | 997.99 (37.45%) | 766.64 (28.77%) | 241.57 (9.07%) | |
| Flood | | 950.52 (35.67%) | 463.98 (17.41%) | 618.29 (23.2%) | 37.25 (1.4%) | 594.87 (22.32%) |

III. Results and discussion

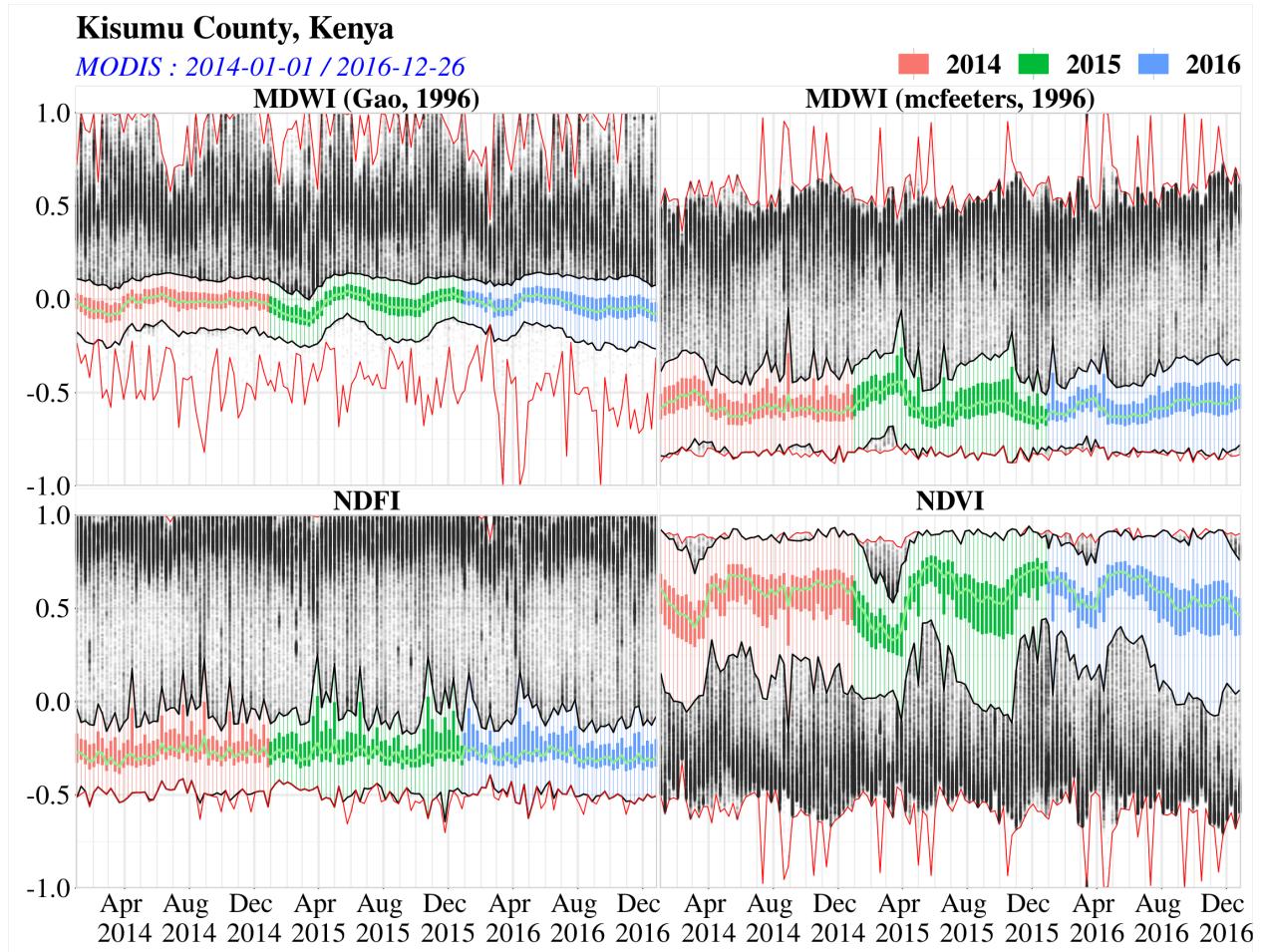


Figure 3: Temporal variability in water and vegetation in Kisumu county, Kenya.

III.1. Prior spatial distribution of the spatial data node states

The results of the processed spatial data nodes used to feed the BNs model are presented in Figure ???. As it can be seen, the topographic characteristics of the area are captured in both slope and flow accumulation nodes. In Kisumu county, run-off flows from East to West with most of the flow converging to Lake Victoria which is the flattest land located in vast low laying lands. In the East of the Lake basin are the steepest lands of Kisumu causing the essential of surface water to flow Westwards. Most poor-quality data are located

Kisumu County, Kenya: MODIS 2014–01–01 / 2016–12–26.

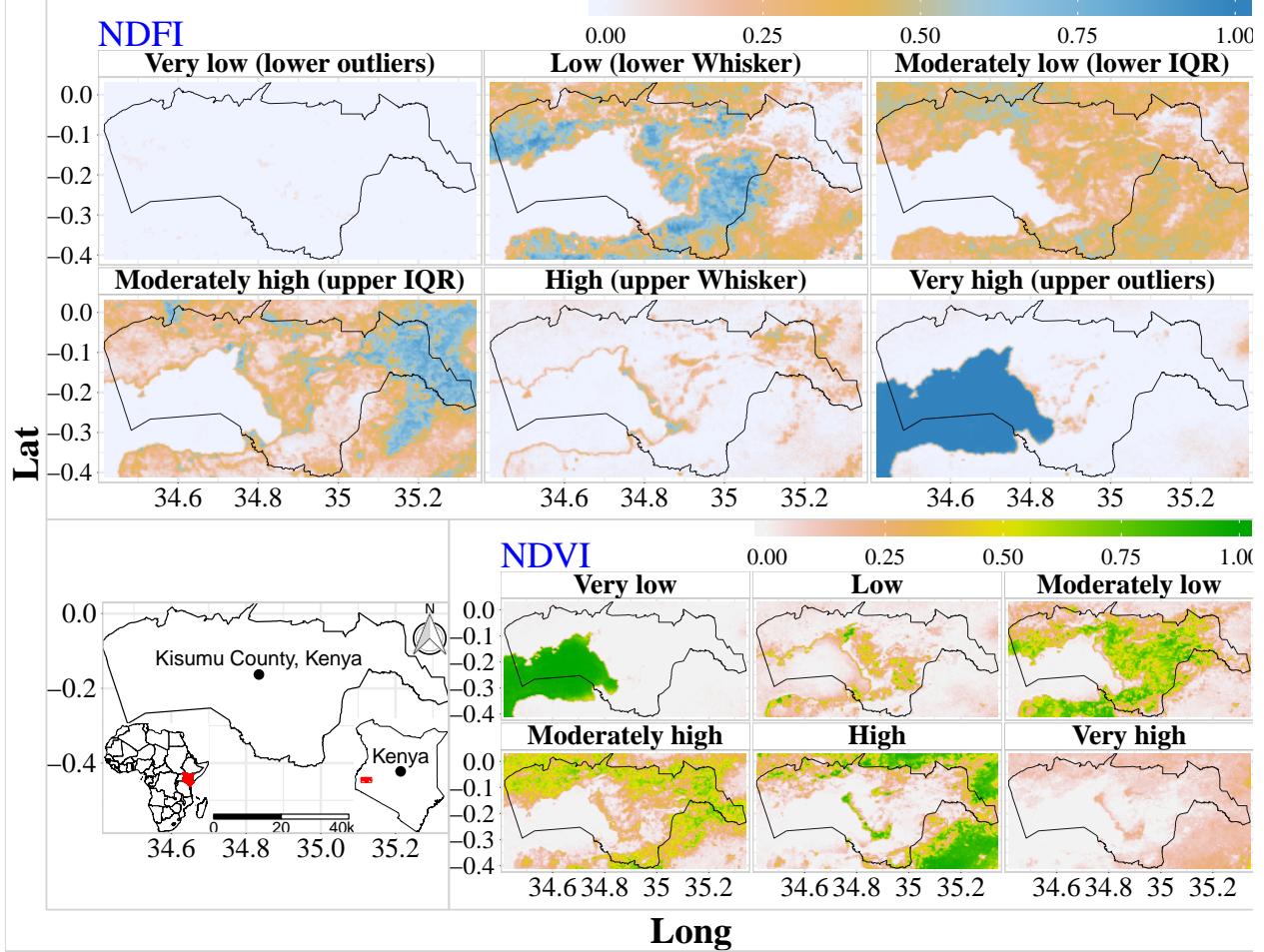


Figure 4: Discrete states of vegetation and water in Kisumu county, Kenya.

around the water bodies where the node power of tools scored the lowest values to an extent that most water bodies are clearly visible. The chance of encountering water on the ground (water present node), either in forms of floods water or water body, is high in most area of Kisumu. This seems to be frequent towards the beginning of the growing season since there is no big difference between the map of the node water present and that of the node flood at the beginning of the rainy season. In general, vegetation in Kisumu experiences high intra-annual variation with, unexpectedly, the highest variability associated with Lake Victoria. 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. As expected, the spatial pattern of soil water content mimic that of the exposure to wetness (also water present and flood at the beginning of the rainy season) with the Lake Victoria scoring the highest values. In general, there a drastic moisture within the flat area beyond the lake water; which moisture increases towards the eastern and the northern parts of the county. The highest sensitivity to flood is encountered in the Lake shore. Some forested areas are wrongly detected as high flood sensitivity spots in the north and west eastern of the county towards the neighboring Nandi and Kericho counties. Vegetation sensitivity to water variation, which seems to be negatively correlated with forest and open water, decrease from the literal Lake Victoria outwards.

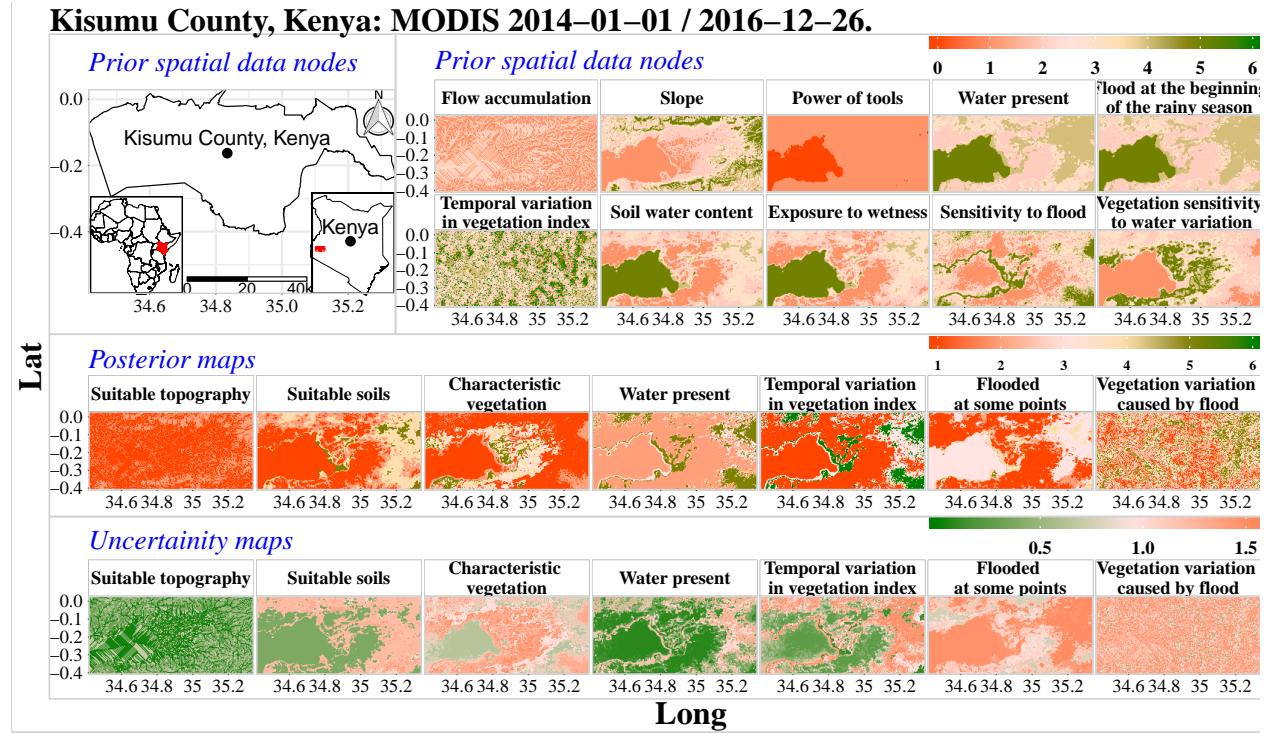


Figure 5: Prior, posterior, and uncertainty maps of different spatial data used to feed the BNs for mapping FBFS in Kisumu county, Kenya.

The importance of vegetation infestation in the Kenyan part of Lake Victoria has been reported in previous studies.⁵³ The high vegetation along with its low sensitivity to water variation in littoral zone can be explained by the permanence of water. Gikuma-Njuru et al.⁵⁴ reported higher Chlorophyll concentrations in the littoral compared to the pelagic zones due to nutrients additions from tributaries.⁵⁵ The highest vegetation variability associated with the lake water may be explained by a sequence of proliferation and decline of aquatic plants (e.g. algae) at certain periods of the year. Another explanation could be the periods of active mixing resulting in water-vegetation mixtures.^{53,54} The confusion with regards to the spectral response of open water and forest is a testimonies of uncertainty sources. Care must be taken when dealing with remote sensing data to avoid such errors, particularly in studies dealing with water and vegetation.⁵³ In general, water indices are sensitive to vegetation liquid water.¹⁰

III.2. Posterior spatial distribution of the spatial data node states

The output maps from the BNs model are shown in Figure ?? and their respective uncertainties in Figure ???. The suitable topography for FBFS, resulting from the interaction between the slope and flow accumulation nodes, is generally high in area having high flow accumulation which are not necessarily water body. The soil suitability (suitable soils node) is even more misleading showing forests and open waters as having highest suitable in soil for FBFS. Intuitively, permanently waterlogged areas may not be suitable for agriculture. Characteristic vegetation, in contrast, is rather accurate excluding most open waters and forests from typical vegetation for FBFS. Interestingly, the pattern of the characteristic vegetation is depicted with most vegetation exposed to waterlogging close to open water. Clearly, there is a change in the characteristic vegetation at a certain distance from Lake Victoria with the closed area to the Lake showing vegetation exposed to waterlogging and most of the areas near rivers Mogus, Awach, Sondu Miriu, Nyando and Kibos exhibiting hydrated vegetation.

Important changes were introduced from the prior to the posterior maps of water present node. One important change is related to the appearance of a sort-of islands or at least a tendency to drying in certain parts of the Lake shores leading to the conclusion that water may be not be present all the yearlong in that area. Another important change is the increase in the areas where water can be expected on the ground. Probably due to vegetation liquid water, most forests have shown important water presence. The variable temporal variation in vegetation index ranked higher on forests and open waters with most central areas resulting in the lowest. This high vegetation variability over the Lake support the idea of the sequential increase / decrease of aquatic vegetation at certain period of the year (see section III.1.), particularly looking at the map of the node vegetation variation caused by flood. The highest chance of flood is recorded near the lake shore and river Nyando, but most floodplains scored high values. As expected, vegetation variation attributed to flood (vegetation variation caused by flood node) seems to be positively correlated with flow accumulation and suitable topography for FBFS. In general, most of the posterior maps are not conclusive enough to be considered as standalone decision support with regards the potential area for FBFS. The target map (pixel is FBFS) seems to also be strongly correlated with flow accumulation and suitable topography and support the idea of vegetation presence inside the Lake which

Table 3: Spatial coverage of FBFS in Kisumu County, Kenya and Tigray region, Ethiopia.

| | Highly unlikely | Unlikely | Possible | Likely | Highly likely |
|-----------------|-------------------|-------------------|-------------------|------------------|-------------------|
| Kisumu 1 | 548.69 (20.59%) | 1085.89 (40.75%) | 27.03 (1.01%) | 217.09 (8.15%) | 786.21 (29.5%) |
| Tigray 1 | 26.81 (0.05%) | 23568.4 (44.85%) | 1028.91 (1.96%) | 9009.07 (17.14%) | 18885.03 (35.94%) |
| Kisumu 2 | 580.62 (21.79%) | 769.4 (28.87%) | 789.19 (29.62%) | 277.32 (10.41%) | 248.38 (9.32%) |
| Tigray 2 | 12943.07 (24.63%) | 13759.91 (26.18%) | 13381.08 (25.46%) | 3421.86 (6.51%) | 9012.3 (17.15%) |

certain parts appears among the very likely areas for FBFS. As expected, most riverbanks are detected as having high potential for FBFS; which potential declines with distance.

The spatial explicit uncertainty (Figure ??), expressed as shannon entropy, revealed that the lowest level of uncertainty seems to be achieved with the suitable topography. This may be explained by the simplicity of the causality specified in this node but also the relative accuracy of the data (radar versus optical sensor) used to feed the parents nodes (slope and flow accumulation). Regarding this node, the highest uncertainty spot seems to be located around high flow accumulation areas where water is potentially present, but the topography is not necessarily flat. As stated earlier, the node suitable soils might be misleading, and this can be seen in the uncertainty map where most areas having either relatively low vegetation or water ranked amongst the highest uncertainty spots. Most open water seems to have the lowest uncertainty with regards to characteristic vegetation node contrary to water presence. The uncertainty map related to temporal variation in vegetation index has captured the drying spot inside the Lake as discussed earlier. High uncertainty regarding flood seems be ubiquitous and this seems to be propagated into the node vegetation variation caused by flood. Coming to the target node, there is high uncertainty on most pixels associated with open water. As it can be seen, the water in the Lake is associated with the highest level of uncertainty supporting the ideas that these areas are not suitable. In a nutshell, the posterior maps are to be used with consideration to the associated uncertainty maps.

III.3. Case study: Application to Tigray region of Ethiopia

III.4. Potential of FBFS in Kisumu and Tigray

Tigray region, Ethiopia: MODIS 2014–01–01 / 2016–12–26.

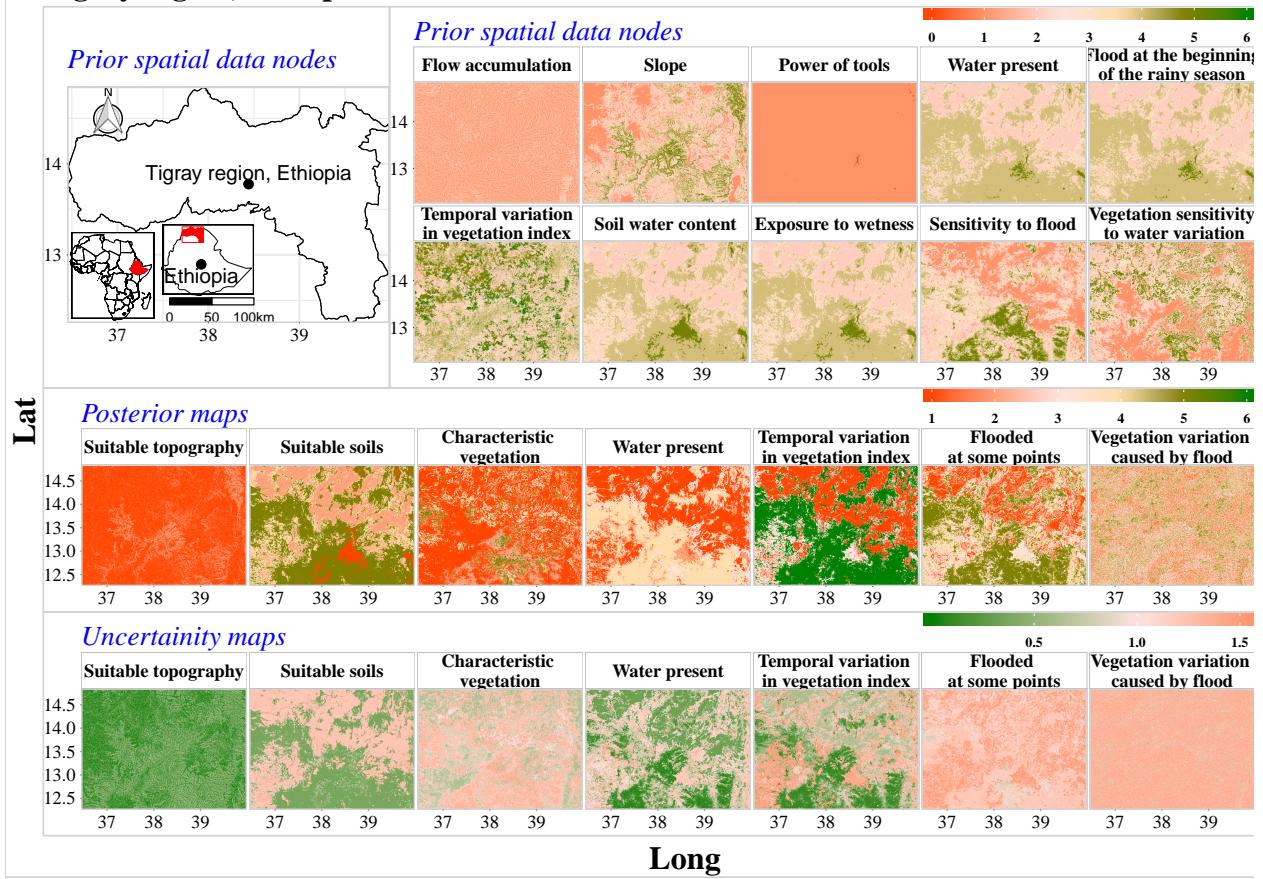


Figure 6: Prior, posterior, and uncertainty maps of different spatial data used to feed the BNs for mapping FBFS in Tigray region, Ethiopia

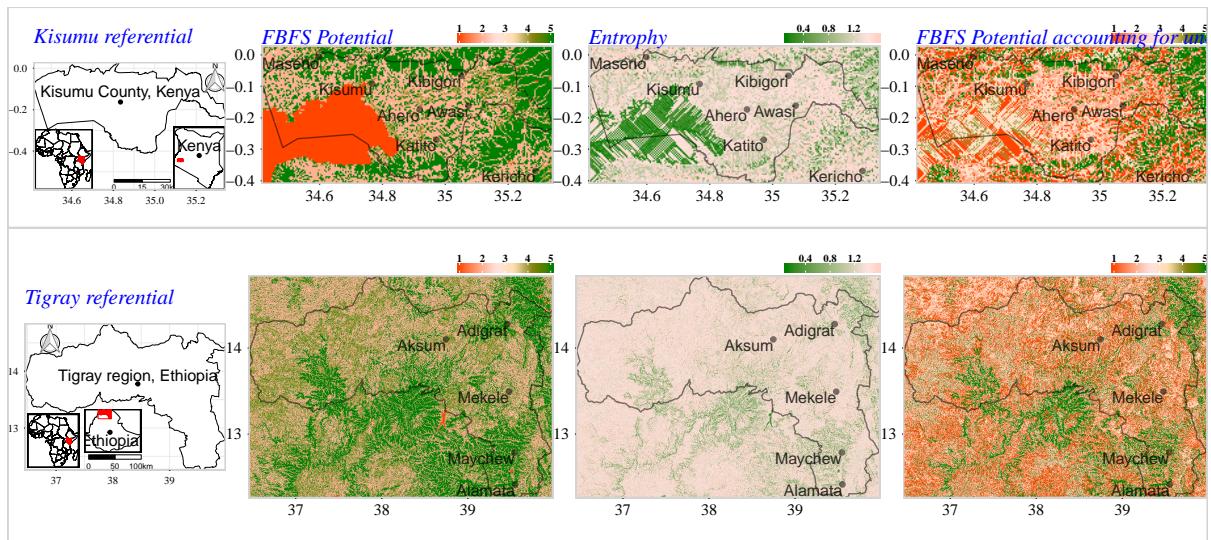


Figure 7: Potential areas for FBFS and associated uncertainty in Kisumu County, Kenya and Tigray region, Ethiopia.

References

1. Puertas, D. G.-L., Steenbergen, F. van, Haile, A. M., Kool, M. & Embaye, T.-a. G. *Flood based farming systems in Africa*. 52 (2011).
2. Steenbergen, F. van, Lawrence, P., Mehari, A., Salman, M. & Faurès, J.-M. *Guidelines on spate irrigation*. 249 (FAO, 2010).
3. Varisco, D. M. Sayl and ghayl: The ecology of water allocation in Yemen. *Human Ecology* **11**, 365–383 (1983).
4. Xing, L. *et al.* China wetland extraction and classification using MODIS data. in *2014 third international workshop on earth observation and remote sensing applications (eorsa)* 349–352 (IEEE, 2014).
5. FBLN. Flood-based Livelihood Network (FBLN) Foundation. (2018).
6. Haile, A. M. *A Tradition in Transition, Water Management Reforms and Indigenous Spate Irrigation Systems in Eritrea*. 212 (CRC Press, 2010).
7. Ghebreamlak, A. Z., Tanakamaru, H., Tada, A., Adam, B. M. & Elamin, K. A. E. Satellite-based mapping of cultivated area in Gash Delta Spate Irrigation System, Sudan. *Remote Sensing* **10**, 186 (2018).
8. Khalid, A. E. *et al.* Performance Evaluation of Spate Irrigation Using Remote Sensing and DEM. in *The 7th international conference on water resources and environment research (icwrer2016)* (ICWRER2016, 2016).
9. Theilen-Willige, B. *et al.* Flash Floods in the Guelmim Area/Southwest Morocco—Use of Remote Sensing and GIS-Tools for the Detection of Flooding-Prone Areas. **5**, 203–221 (2015).
10. 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).
11. Wegmann, M., Leutner, B. & Dech, S. *Remote Sensing and GIS for Ecologists: Using Open Source Software (Data in the Wild)*. 324 (Pelagic Publishing, 2016).
12. Konecny, G. *Geoinformation: remote sensing, photogrammetry and geographic information systems*. 472 (CRC Press/Taylor & Francis Group, 2014).
13. Bashari, H., Smith, C. & Bosch, O. J. H. Developing decision support tools for range-land management by combining state and transition models and Bayesian belief networks. *Agricultural Systems* **99**, 23–34 (2008).
14. Hubbard, D. W. *How to Measure Anything: Finding the Value of Intangibles in Business*. 432 (John Wiley & Sons, Inc., 2014).
15. Kuhnert, P. M., Martin, T. G. & Griffiths, S. P. A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters* **13**, 900–914 (2010).

16. Kuhnert, P. M., Martin, T. G., Mengersen, K. & Possingham, H. P. Assessing the impacts of grazing levels on bird density in woodland habitat: A Bayesian approach using expert opinion. *Environmetrics* **16**, 717–747 (2005).
17. Whitney, C., Shepherd, K. & Luedeling, E. *Decision analysis methods guide; Agricultural policy for nutrition.* 27 (2018). doi:[10.5716/WP18001.PDF](https://doi.org/10.5716/WP18001.PDF)
18. Kuhnert, P. M. Four case studies in using expert opinion to inform priors. *Environmetrics* **22**, 662–674 (2011).
19. Luedeling, E. *et al.* Fresh groundwater for Wajir - ex-ante assessment of uncertain benefits for multiple stakeholders in a water supply project in Northern Kenya. *Frontiers in Environmental Science* **3**, (2015).
20. Whitney, C. *et al.* Probabilistic Decision Tools for Determining Impacts of Agricultural Development Policy on Household Nutrition. *Earth's Future* **6**, 359–372 (2018).
21. Yet, B. *et al.* A Bayesian Network Framework for Project Cost, Benefit and Risk Analysis with an Agricultural Development Case Study. *Expert Systems with Applications* **60**, 141–155 (2016).
22. Pourret, O., Naïm, P. & Marcot, B. *Bayesian Networks: A Practical Guide to Applications.* 446 (John Wiley & Sons Ltd, 2008).
23. Steenbergen, F. van, MacAnderson, I. & Mehari, A. *Spate irrigation in the Horn of Africa: Status and potential.* 44 (2011).
24. Arge, L. *et al.* Efficient flow computation on massive grid terrain datasets. *GeoInformatica* **7**, 283–313 (2003).
25. 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).
26. 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).
27. 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).
28. Ji, L., Zhang, L. & Wylie, B. Analysis of Dynamic Thresholds for the Normalized Difference Water Index. **75**, 1307–1317 (2009).
29. 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).
30. 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).
31. Roy, D. *et al.* Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment* **145**, 154–172 (2014).

32. Tarboton, D. Terrain analysis using digital elevation models in hydrology. in *23rd esri international users conference* (2003).
33. Tucker, C. J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment* **08**, 127–150 (1979).
34. 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).
35. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2018).
36. Kisumu County Government. *Kisumu County First Integrated Development Plan 2013 - 2017*. (2013).
37. Farr, T. G. & Kobrick, M. Shuttle radar topography mission produces a wealth of data. *Eos* **81**, 583–585 (2000).
38. Farr, T. G. *et al.* The shuttle radar topography mission. *Reviews of Geophysics* **45**, (2007).
39. Nikolakopoulos, K. G., Kamaratakis, E. K. & Chrysoulakis, N. SRTM vs ASTER elevation products. Comparison for two regions in Crete, Greece. *International Journal of Remote Sensing* **27**, 4819–4838 (2006).
40. Jarvis, A., Reuter, H. I. I., Nelson, A. & Guevara, E. Hole-filled seamless SRTM data V4. (2008).
41. Falorni, G., Teles, V., Vivoni, E. R., Bras, R. L. & Amaralunga, K. S. Analysis and characterization of the vertical accuracy of digital elevation models from the Shuttle Radar Topography Mission. *Journal of Geophysical Research: Earth Surface* **110**, (2005).
42. King, M. D., Herring, D. D. & Diner, D. J. The earth observing system (EOS): A space-based program for assessing mankind's impact on the global environment. *Opt. Photon. News* **6**, 34–39 (1995).
43. Vermote, E. MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006. (2015).
44. Pearl, J. *Causality: Models, Reasoning and Inference*. 1–386 (Cambridge University Press, 2000). doi:[citeulike-article-id:3888442](#)
45. GADM. Database of Global Administrative Areas. (2018).
46. Hijmans, R. J. raster: Geographic Data Analysis and Modeling. (2019).
47. Mattiuzzi, M. & Detsch, F. MODIS: Acquisition and Processing of MODIS Products. (2018).
48. Busetto, L. & Ranghetti, L. MODIStsp: An R package for automatic preprocessing of MODIS Land Products time series. *Computers & Geosciences* **97**, 40–48 (2016).
49. contributors, G. GDAL/OGR Geospatial Data Abstraction software Library. (2018).

50. Likert, R. A technique for the measurement of attitudes. *Archives of Psychology* **22**, 5–55 (1932).
51. Wickham, H. & Stryjewski, L. *40 Years of Boxplots*. 1–17 (2012).
52. Hothorn, T., Hornik, K. & Zeileis, A. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* **15**, 651–674 (2006).
53. Cavalli, R. M., Laneve, G., Fusilli, L., Pignatti, S. & Santini, F. Remote sensing water observation for supporting Lake Victoria weed management. *Journal of Environmental Management* **90**, 2199–2211 (2009).
54. Gikuma-Njuru, P. *et al.* Eutrophication of the Lake Victoria Ecosystem. in *Lake victoria regional water quality synthesis report* (eds. Mwanuzi, F. L., Abuodha, J. O. Z., Muyodi, F. J. & Hecky, R. E.) 81 (LVEMP, 2005).
55. World Bank. *Kenya, Tanzania, Uganda: Lake Victoria Environmental Management Project*. 158 (World Bank, 1996).