Mapping Flood-based Farming Systems with Bayesian networks.

Issoufou, Liman Harou[a,b,], Cory Whitney[c], James Kungu[a], Eike Luedeling[c]

[a]*Kenyatta University, Department of Environmental Sciences, P.O. Box 43844 00100 Nairobi, Kenya*
[b]*World Agroforestry Centre (ICRAF), United Nations Avenue, Gigiri, P.O. Box 30677-00100, Nairobi, Kenya*
[c]*University of Bonn, Department of Horticultural Sciences, Auf dem Hügel 6, D-53121, Bonn, Germany*
[d]*Center for Development research (ZEF), University of Bonn, Genscherallee 3, D-53113, Bonn, Germany*

Abstract

Many actors in agricultural research, development and policy require accurate information on the spatial extents of cropping and farming practices. While remote sensing provides venues for obtaining such information, it is often difficult to distinguish between agricultural practices which can be quite similar in their spectral signatures. This can easily lead to misclassification by traditional algorithms. We addressed this limitation by combining traditional remote sensing techniques with probabilistic reasoning engines informed by expert knowledge to map flood-based farming systems, an important form of agricultural production employed by farmers in regions with seasonal water surplus. Flood-based farming systems were mapped across Kisumu County in Kenya and the Tigray region in Ethiopia using a model of causality. While the two areas present noticeable differences in terms of their hydrology, vegetation, or the practice of agronomic flooding, these relevant aspects for mapping flood-based farming systems are mostly stochastic in space and time, which makes it difficult to provide a generalizable characterization of the practice using precise quantifiers. Recognizing this aspect, the causal model described in this paper considers

Email addresses: issoufoul@gmail.com; L.issoufou@cgiar.org (Issoufou, Liman Harou), cory.whitney@uni-bonn.de (Cory Whitney), kungu.james@ku.ac.ke (James Kungu), luedeling@uni-bonn.de (Eike Luedeling)

uncertainties across expert-based estimates to provide quantitative assessment of the overall uncertainty in the final output. To account for the effect of specific geographic settings, we used Bayesian networks informed by expert knowledge to characterize these settings. Three years (2014-2016) of processed MODIS VI data were acquired in the forms of normalized difference spectral indices which were used as proxies for estimating various spatial metrics relevant for flood-based farming systems detection. The spatial metrics along with other non-spatial data derived from expert judgement were used as nodes to feed the Bayesian networks. Results were transparently generated in the form 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 spatially explicit uncertainties. We demonstrate that spatially explicit information can be derived from remote sensing data as fuzzy linguistic quantifiers which are suitable for representing node states in Bayesian networks. When such metrics are available, Bayesian networks 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 achieved a remarkably accurate result in both settings where 60 to 100% of various flood-based farming system fields sampled from different locations were correctly mapped with high chance of being suitable for the practice.

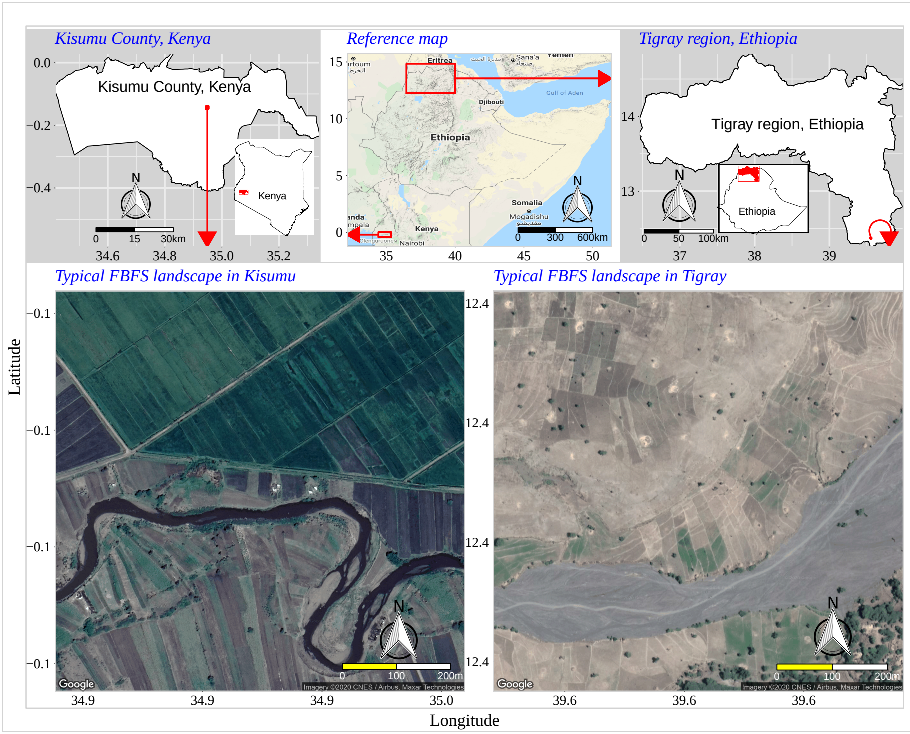

Figure 1: Typical flood-based farming systems landscapes found in Tigray region of Ethiopia and Kisumu County in Kenya.

Introduction

1. Study area

2. Material and methods

2.1. Conceptual framework

2.2. Sensor Characteristics

2.3. Data acquisition and pre-processing

2.3.1. The BNs

2.3.2. Spatial data

2.4. Making sense of data in specific contexts

2.5. Deriving spatial data nodes

2.5.1. General multi-layer procedure applied to time series data

dddddddddddddddddddd hhhhhhhhhhhhhhhhhhhhhhh

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

for *each layer in the time series* **do**

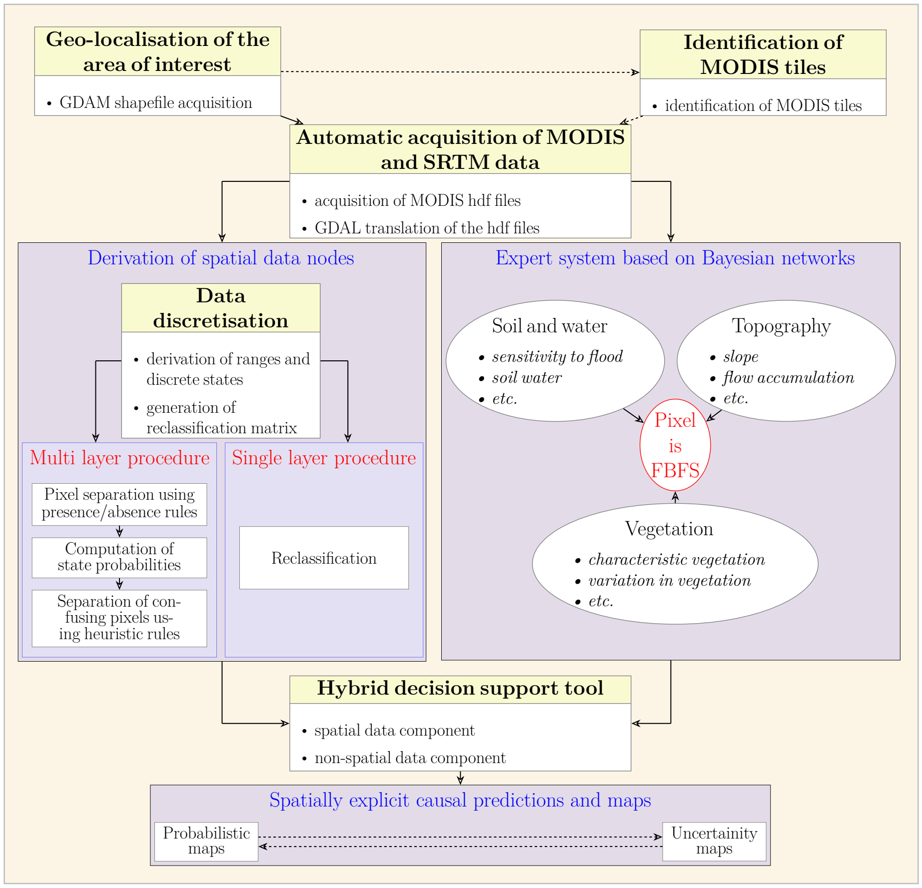

Figure 2: Conceptual framework for mapping FBFS in Kisumu County, Kenya and Tigray region, Ethiopia.

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

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).

nnnnnnnnnn nnnnnnnnn nn nnnnnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnn

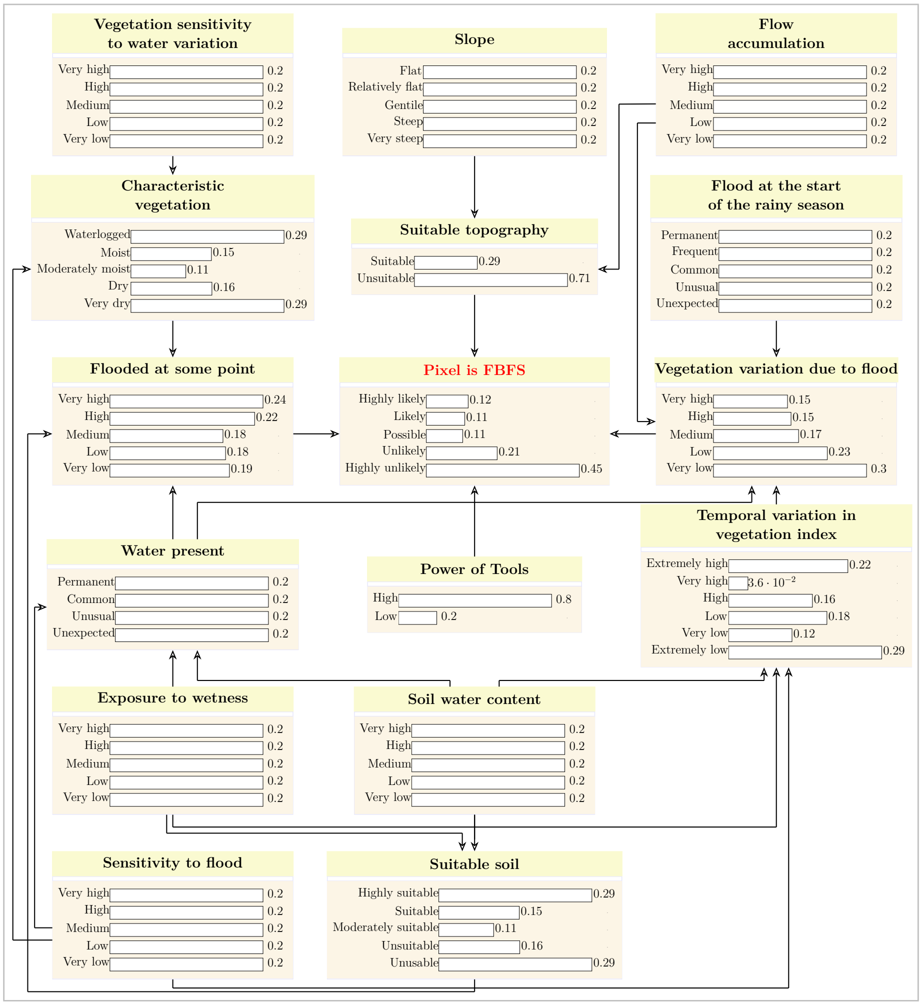

Figure 3: Bayesian Network describing the causal reasoning used for mapping FBFS in Kisumu, Kenya.

nnnnnnnn n n nnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

nnnnnnnnnn nnnnnnnnn nn nnnnnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnn nnnnnnnn n n nnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

nnnnnnnnnn nnnnnnnnn nn nnnnnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnn

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)

nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnn

nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnn

nnnnnnnn n n nnnnnnnnnnnnnnnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

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 maximum (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 maximum 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
```

2.5.2. General procedure applied to single-layer data

2.5.3. Specific procedure used for spatial data generation

3. Results

3.1. Spatio-temporal analysis of water and vegetation in Kisumu County, Kenya

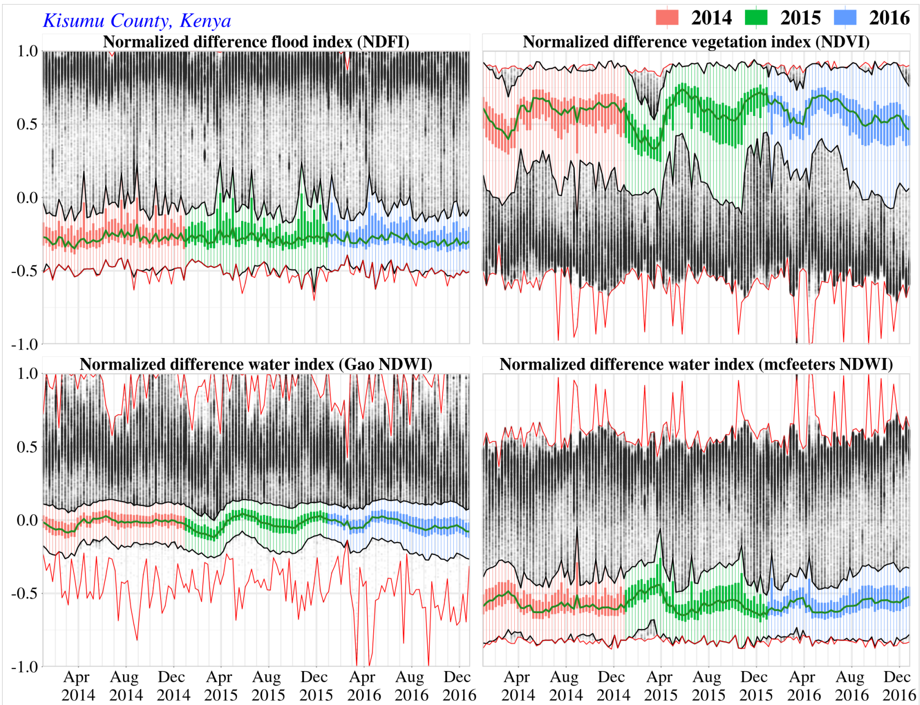

Figure 4: Temporal variability in water and vegetation in Kisumu county, Kenya.

Table 2: Spatial coverage of FBFS in Kisumu County, Kenya and Tigray region, Ethiopia.

	Highly unlikely	**Unlikely**	**Possible**	**Likely**	**Highly likely**
Kisumu	379.7 (14.2%)	1033.1 (38.8%)	30.2 (1.1%)	285.4 (10.7%)	936.5 (35.1%)
Tigray	20.8 (0.0%)	22727.9 (43.2%)	831.3 (1.6%)	6710.1 (12.8%)	22228.1 (42.3%)

Table 3: Prediction accuracy of a spatially explicit bayesian network used for mapping FBFS in Kisumu County, and Tigray region, Ethiopia.

Study area	**Validation sample**	**Highly unlikely**	**Unlikely**	**Possible**	**Likely**	**Highly likely**
Kisumu	Bare lands	66.7%	33.3%	0.0%	0.0%	0.0%
	FBFS Fields	0.0%	0.0%	0.0%	0.0%	100.0%
	Forests	0.0%	100.0%	0.0%	0.0%	0.0%
	Irrigated agricultural fields	0.0%	20.5%	2.3%	2.3%	75.0%
	Rainfed agricultural fields	0.0%	80.0%	0.0%	0.0%	20.0%
	Riperian forests	0.0%	0.0%	0.0%	0.0%	100.0%
	Settlements	0.0%	63.6%	0.0%	27.3%	9.1%
	Vegetated plateaux	0.0%	88.9%	0.0%	0.0%	11.1%
	Water bodies	65.0%	7.1%	0.0%	27.9%	0.1%
Tigray	Bare lands	0.0%	62.3%	0.6%	2.6%	34.4%
	FBFS fields	0.0%	24.5%	0.0%	1.9%	73.6%
	Forests	0.0%	77.8%	0.0%	11.1%	11.1%
	Irrigated agricultural fields	0.0%	65.4%	0.0%	0.0%	34.6%
	Rainfed agricultural fields	0.0%	100.0%	0.0%	0.0%	0.0%
	Riperian forests	0.0%	30.6%	2.8%	5.6%	61.1%
	Settlements	0.0%	60.0%	0.0%	0.0%	40.0%
	Vegetated plateaux	0.0%	16.6%	0.0%	50.1%	33.3%
	Water bodies	78.2%	4.9%	0.0%	16.9%	0.0%

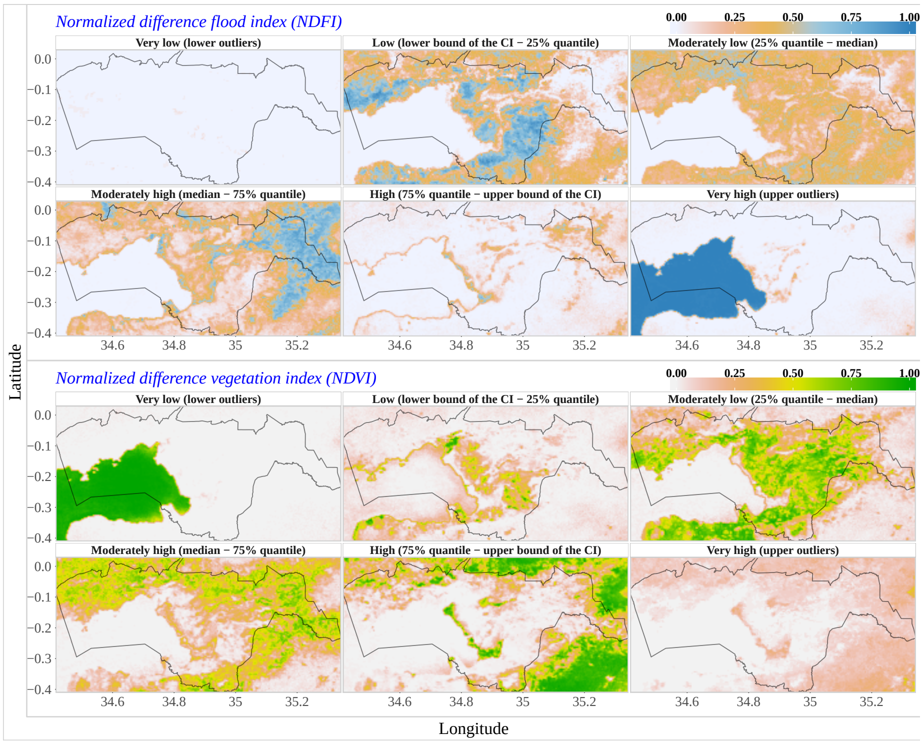

Figure 5: Discrete states of vegetation and water in Kisumu county, Kenya.

3.2. FBFS-relevant metrics

3.2.1. Prior distribution of FBFS-relevant metrics

3.2.2. Posterior distribution of FBFS-relevant metrics

3.2.3. Uncertainty of FBFS-relevant metrics

3.3. FBFS potential in Kisumu County, Kenya and Tigray region, Ethiopia

3.4. Accuracy assessment

3.5. Uncertainty-based predictions

4. Discussion

5. Conclusion

Acknowledgements

This study was jointly funded by the Deutscher Akademischer Austauschdienst (German Academic Exchange Service; DAAD) and the World Agroforestry

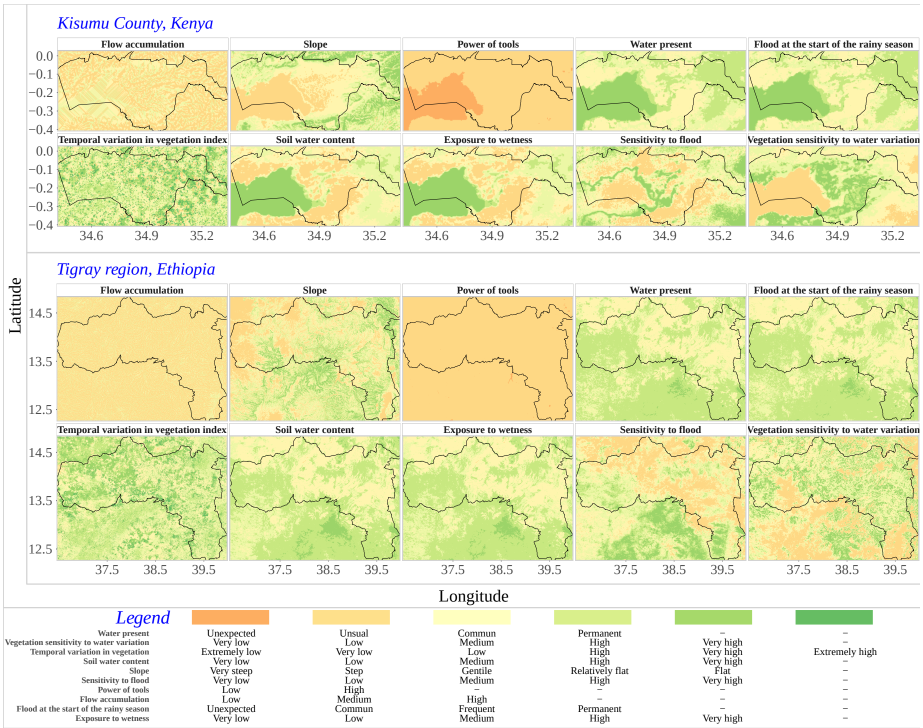

Figure 6: Prior maps of differents spatial data used to feed the BNs for mapping FBFS in Kisumu county, Kenya and Tigray region, Ethiopia.

Centre (International Centre for Research in Agroforestry; ICRAF). We thank the farmers and experts who contributed in various ways during the development of the model. We thank the department of Dryland Agriculture of Mekelle University for guidance on the sampling frame in the Tigray region and for facilitating logistics during field work. Thanks to the participants of the leadership course in flood-based farming and water harvesting in Kenya and the participants of the International Training on Integrated Watershed Management and FBFS in Ethiopia, who drafted the models.

References

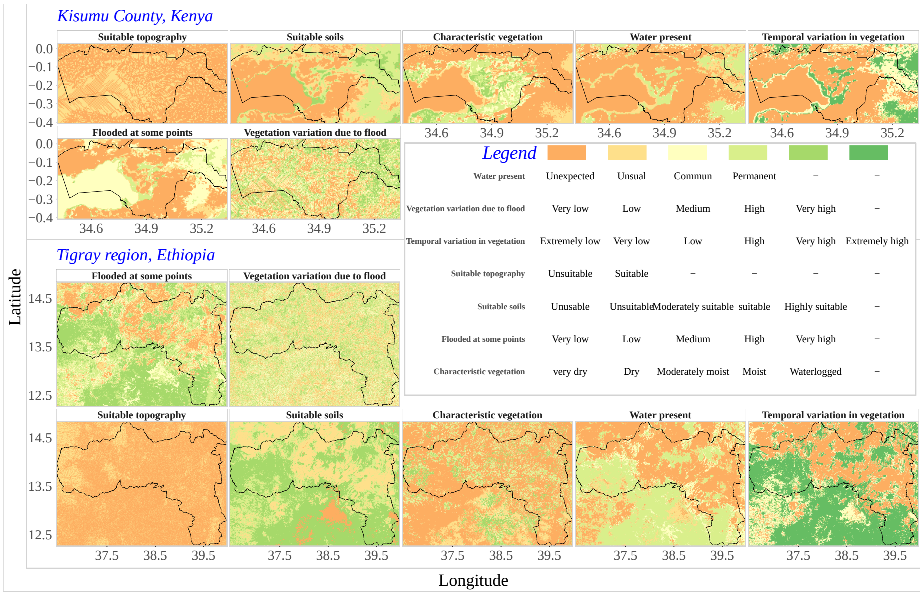

Figure 7: Posterior maps of differents spatial data used to feed the BNs for mapping FBFS in Kisumu county, Kenya and Tigray region, Ethiopia.

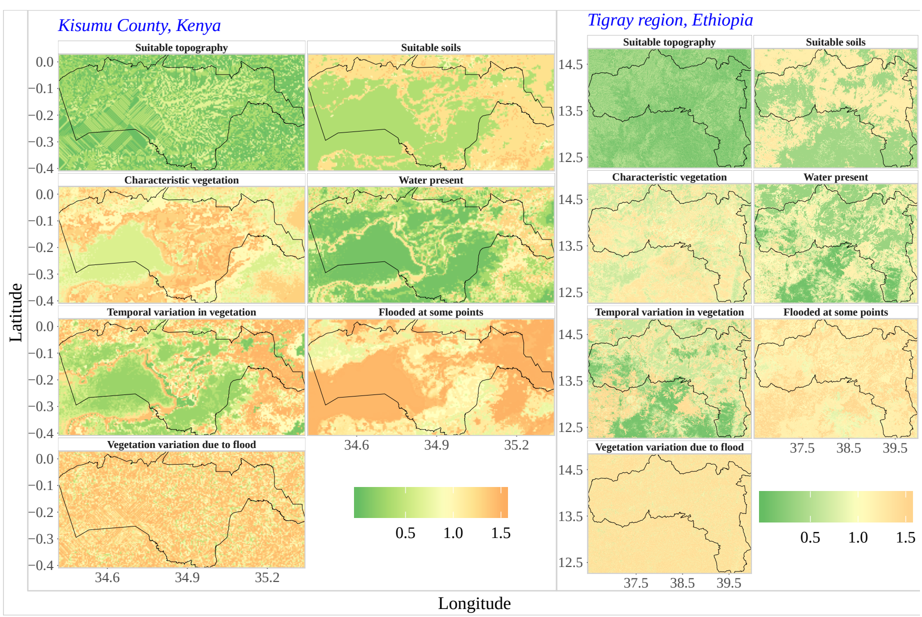

Figure 8: Uncertainty maps of differents spatial data used to feed the BNs for mapping FBFS in Kisumu county, Kenya and Tigray region, Ethiopia.

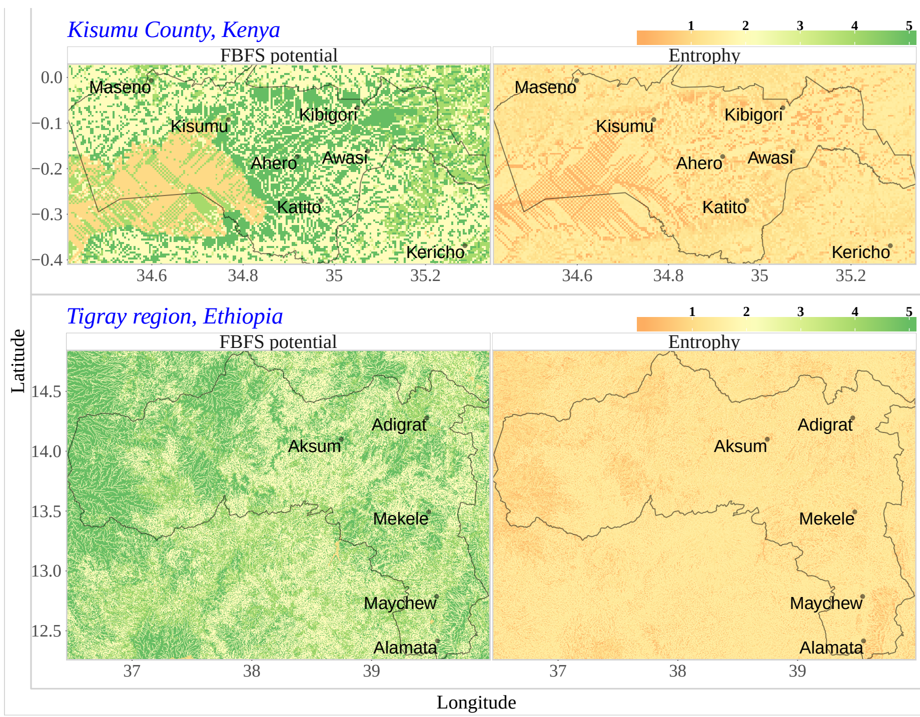

Figure 9: Potential areas for FBFS and associated uncertainity in Kisumu County, Kenya and Tigray region, Ethiopia.

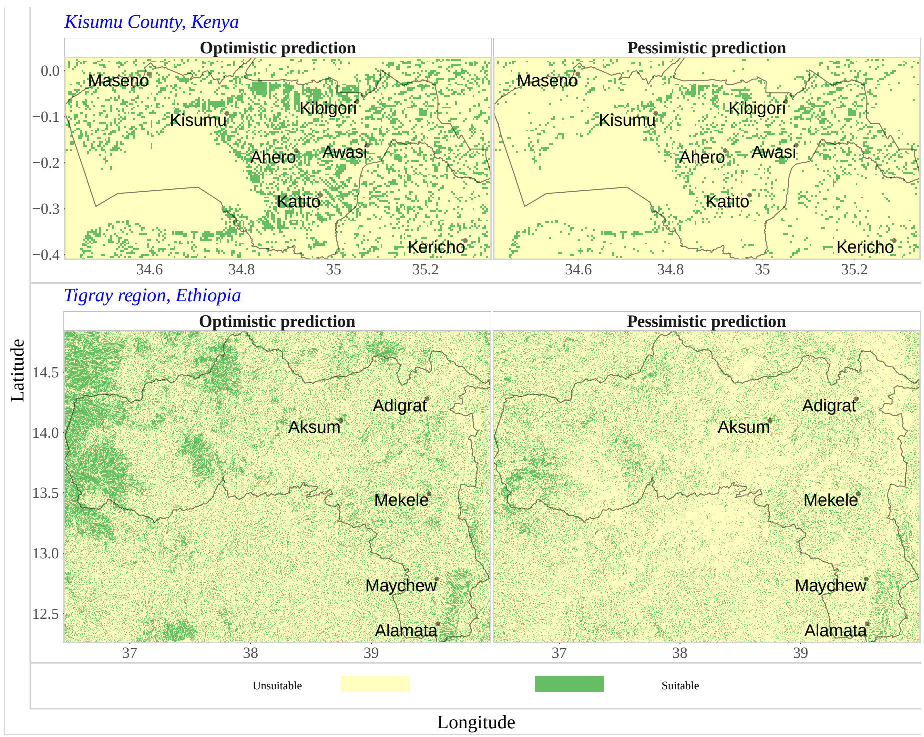

Figure 10: Potential areas for FBFS based different level of uncertainities in Kisumu County, Kenya and Tigray region, Ethiopia.