Applied Geography 96 (2018) 86–97

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

# Predictors determining the potential of inland valleys for rice production development in West Africa

Justin Fagnombo Djagba[a,b,*], Luc O. Sintondji[c], Amadou Malé Kouyaté[d], Idriss Baggie[e], Georges Agbahungba[b], Abdoulaye Hamadoun[d], Sander J. Zwart[f]

[a] Africa Rice Center (AfricaRice), 01 BP 2031 Cotonou, Benin
[b] International Chair in Mathematical Physics and Applications (ICMPA – UNESCO Chair), University of Abomey-Calavi, 072 BP 50 Cotonou, Benin
[c] Hydraulic and Water Control Laboratory, National Institute of Water, University of Abomey-Calavi, 01 BP 526 Cotonou, Benin
[d] Regional Agricultural Research Center of Sikasso, Institut d'économie rurale (IER), BP 16 Sikasso, Mali
[e] Sierra Leone Agricultural Research Institute (SLARI), Tower Hill, P.M.B. 1313 Freetown, Sierra Leone
[f] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE, Enschede, Netherlands

ARTICLE INFO

*Keywords:*
Inland valley
Rice production
Predictor
Relevant variable selection
Random forests

ABSTRACT

Water availability and high soil fertility make inland valley landscapes suitable for sustainable rice-based cropping. In this study, Random Forests statistical analysis was used on a database of 499 surveyed inland valleys in four study zones in three West African countries. The goal of the study was to assess parameters that indicate (are predictors for) high potential for development of rice-based systems in inland valleys. These parameters are related to the biophysical (hydrology, soil, climate, and topography) and socio-economic (demography, accessibility, and markets) environments. Farmer group surveys and secondary data from existing publicly available spatial data sets were used.

The analysis revealed that, across the four research areas, the following parameters were relevant predictors for rice development: (1) distance from the inland valley to the nearest market; (2) distance from the inland valley to the nearest rice mill; (3) population density in the immediate environment of the inland valley; (4) total nitrogen in the top 20 cm of the soil profile; (5) land elevation; and (6) soil texture on the upper slope of the inland valley. Several predictors were highly important for specific research areas, but not for all, thus showing the diversity in the studied agricultural landscapes. These predictors included soil fertility management, source of irrigation water, and the percentage of female farmers in the inland valley. The identified relevant predictors will be used to map the potential rice production development of the inland valleys. This will help development agencies to assess their zones based on quantitative analysis for inland valley potential development.

## 1. Introduction

Crop yields are generally poor in West Africa; crop production is insufficient and West Africa depends on food imports (Niang et al., 2017; Seck, Diagne, Mohanty, & Wopereis, 2012). Thus, food insecurity is a major problem. It was estimated that there were 239 million hungry people in sub-Saharan Africa (SSA) in 2010 (Meijer, Catacutan, Ajayi, Sileshi, & Nieuwenhuis, 2015; Xie, You, Wielgosz, & Ringler, 2014). West Africa has some of the most severe hunger in the world (Brown, Hintermann, & Higgins, 2009; Sasson, 2012). Consequently, West Africa remains a major food buyer, importing large quantities of rice (AfricaRice, 2014) and even local food staples such as millet and maize (Brown et al., 2009). To mitigate food insecurity in this region would require better use of resources, for example, by promoting agricultural use of inland valleys (IVs). These agro-ecosystems conserve moisture and have good soil fertility – good agricultural resources in the face of increasing drought induced by climate change (Van Oort & Zwart, 2018). Various studies (e.g. (Obalum, Nwite, Oppong, Igwe, & Wakatsuki, 2011; Rodenburg et al., 2014; Seck, Tollens, Wopereis, Diagne, & Bamba, 2010; Windmeijer and Andriesse, 1993)) have revealed that West African countries have large untapped IV resources that could be used for rice development. In the face of climatic variability and the effects of climate change, IVs are potential 'bread basket' areas in SSA. The IVs are the main rice cultivation agro-ecosystems in developing countries of SSA (Dossou-Yovo, Baggie, Djagba, & zwart, 2017; McCartney & Houghton-Carr, 2009; Rodenburg et al., 2014; Seck et al., 2012; Worou, Gaiser, Saito, Goldbach, & Ewert, 2012).

Lowland rice cultivation is mainly rainfed and requires stable water

* Corresponding author. Africa Rice Center (AfricaRice), 01 BP 2031 Cotonou, Benin.
*E-mail address:* j.djagba@cgiar.org (J.F. Djagba).

https://doi.org/10.1016/j.apgeog.2018.05.003
Received 16 August 2017; Received in revised form 24 April 2018; Accepted 10 May 2018
Available online 26 May 2018
0143-6228/ 

supply. Thus, large-scale irrigation developments have been promoted on a large scale in West Africa, because of the irrigation potential (Gruber, Kloss, & Shopp, 2009; Musa, 2009; Worou et al., 2012). Unfortunately, many of these implemented irrigation schemes in West Africa have failed or perform below their potentials. (Djagba, Rodenburg, Zwart, Houndagba, & Kiepe, 2014; Inocencio et al., 2007; Nwite, Obalum, Igwe, & Wakatsuki, 2016; Obalum et al., 2011). IVs are the principal rice cultivation agro-ecosystems and the identification of suitable IVs for future rice development is highly important. To know which areas are suitable for rice cultivation requires knowledge of the conditions that favor sustainable rice production development, while limiting environmental impacts (Danvi, Giertz, Zwart, & Diekkrüger, 2018; Djagba, Zwart, Houssou, Tenté, & Kiepe, 2018b). Study of these conditions enables to identify many environmental factors (hydrological, soil, topographical, climatic), socio-economic parameters (extension services, population density, accessibility to road, market, settlement), and farm management practices (e.g. chemical fertilizers, farm technologies) that are important (Abe, Buri, Issaka, Kiepe, & Wakatsuki, 2010; Danvi, Jütten, giertz, Zwart, & Diekkruger, 2016; Gumma, Thenkabail, Fujii, & Namara, 2009; Laborte, Maunahan, & Hijmans, 2012; Masoud, Agyare, Forkuor, Namara, & Ofori, 2013; Qin & Zhang, 2016; Rodenburg et al., 2014). Given the diversity of parameters and factors (variables) that may influence the potential for rice production development in IVs, the most appropriate method for selecting the most important variables must be used. Not all of these parameters and factors contribute to IV agricultural potential to the same degree – some parameters and factors could be more suitable than others. For example, descriptive methods for mapping IVs with agricultural production potential based on expert knowledge were developed and applied (Gumma et al., 2009), but are subjective and may include or exclude parameters or define underestimate or overestimate the importance of a parameter.

The 'Random Forests' model is an approach to map in an unbiased manner the parameters and their importance that contribute or explain a variable (Cutler et al., 2007; Díaz-Uriarte & Alvarez de Andrés, 2006; Hapfelmeier & Ulm, 2013). Random Forests was used by the International Rice Research Institute (IRRI) to map potential rice areas in Laos with a view to limiting environmental degradation due to rice production (Laborte et al., 2012). Elsewhere, Random Forests has been shown to give good accuracy without overfitting and it is relatively robust to outliers and noise (Breiman, 2001; Gislason, Benediktsson, & Sveinsson, 2006; Prasad, Iverson, & Liaw, 2006). Random Forests, considered for classification of multi-source geographic data, presents a comprehensive methodology to assess and analyze classification uncertainty based on the local probabilities of class membership (Gislason et al., 2006; Loosvelt et al., 2012).

The overall aim of this paper was therefore to explore the Random Forests approach to define the best predictors of rice production development in IVs in the diverse landscape of the West African context. An improved understanding of the relevant parameters supports national government agencies, donors and developers in the selection specific IVs or regions and thus increase the chance of success of agricultural development interventions. The specific objectives of this study were: (1) to identify the relevant factors or parameters that define an IV's potential for rice development; and (2) to select the most important parameters as predictive variables, which will then be used to map the potential of IVs for rice production.

The methodological approach, including the Random Forests method that was used in this study is presented in section 2, after the presentation of study area and explanation of heuristics for obtaining the candidate predictors. In section 3, the results on the effectiveness of variable importance measures and the selection of the most relevant predictors among the large number of candidate predictors for the four study areas are reported. In section 4, the methodological approach and the results are discussed, and finally conclusions are presented. The appendix provides the exhaustive list of candidate predictors.

## 2. Materials and methods

### 2.1. Study areas and sampling of inland valleys

This study was carried out in four regions located in Benin, Mali, and Sierra Leone (Fig. 1). A geo-located database was built covering a total of 499 IVs distributed in the four targeted study areas, with 100, 149, 100, and 150 IVs in Mono and Couffo departments (Benin), Ouémé River upper catchment (Benin), Sikasso and Kadiolo circles (Mali), and Bo and Kenema districts (Sierra Leone), respectively.

The selection of IVs was based on many criteria. The first was to locate areas where IV agro-ecosystems are most numerous in West Africa, as in the countries targeted for this study. The specific study regions were selected on the basis of available databases on IVs per country. For Benin, the databases of IMPETUS (a German research project, 2005–2010) and RAP-IV project (Realizing the agricultural potential of IV lowlands in sub-Saharan Africa while maintaining their environmental services, based in Africa Rice Center, 2009–2014) covered the upper of Ouémé River catchment and Mono and Couffo *départements*, respectively (Sintondji et al., 2016). These projects investigated the potential of IVs in target areas. In Mali and in Sierra Leone, the national agricultural research systems (NARS) – Institut d'économie rurale (IER) and Sierra Leone Agricultural Research Institute (SLARI) – have available information on the potential of IVs, their location, and general characteristics (Dossou-Yovo et al., 2017).

The identification of IVs systems was carried out by survey of leaders and key informants at village level. IV location was determined using GPS on the ground or Google Earth to map the boundary. Field surveys were also carried out at IV scale. Another criterion was the spatial distribution of IVs in the target areas. For the established sample of IVs, agricultural use, use for paddy, and crop diversification in IVs were also considered (Table 1).

### 2.2. Candidate predictors

The binomial variable 'presence or absence of rice cultivation in IVs' was used as the dependent variable in this study. Many factors and parameters, defined as variables ($p$), could explain the suitability of an IV for rice production. The project aim was to identify the relevant conditions that suggest rice production potential in IVs and to map the IVs most suitable for rice cultivation development. Using an empirical technique based on expert knowledge and literature review (e.g. (Gumma et al., 2009; Laborte et al., 2012; Masoud et al., 2013; Sakané et al., 2011), a total of 64 variables was proposed. These variables were highly diverse and from many sources (see Appendix). Environmental variables covered hydrological, topographical, climatic, and soil factors. Socio-economic variables related to accessibility and, demographic factors, and IV-use data. Many were also related to farm management practices.

In this study, candidate predictors were identified via three routes: (1) geographical location of IVs and rice sector development elements such as rice mills, markets, and agricultural input stores, and digitizing boundaries of IVs; (2) farmer surveys in IVs to collect environmental, socio-economic, and agricultural use data; and (3) spatial data related to topography, hydrology, climate, soil, accessibility, and population density. In this sub-section, the candidate predictors of rice cultivation potential in IVs are identified and the strategies which could enable their collection in the field or their derivation them from metadata are defined (Fig. 2).

#### 2.2.1. Candidate predictors of location and accessibility

The first step was to identify prospective IVs in the study areas while geo-locating them using GPS. Markets, rice mills, and input stores were also located in the study areas and geographic coordinates were taken. Downloaded, extracted, and symbolized Open Street Map (OSM) data provided much of the location and accessibility data. From OSM nodes,

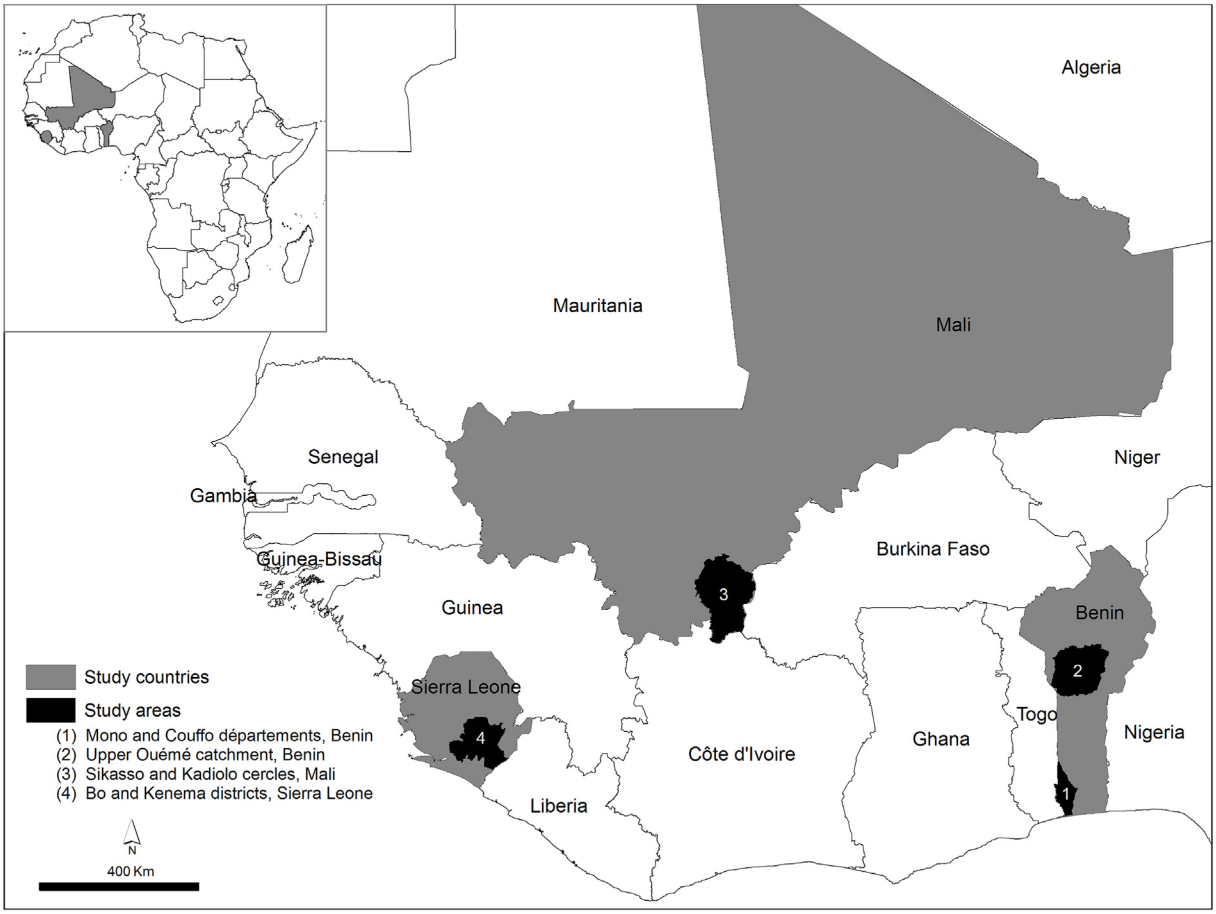


**Fig. 1.** Location of the four study areas: Mono and Couffo *departments*, Benin (1), Upper Ouémé catchment, Benin (2), Sikasso and Kadiolo circles, Mali (3), and Bo and Kenema districts, Sierra Leone (4).

locations of villages, hamlets, towns, and all other settlements such as buildings and residential areas, and some markets were exported. From OSM nodes, all available roads – primary, secondary, tertiary, residential, service, paths, and tracks – were exported and classified. These accessibility and location data were not complete. Digitizing of other roads and other settlements not available in OSM was done manually in Google Earth. To determine the accessibility predictive variables, the numeric values such as the nearest distances from IV to major road (primary road), to other road (all other categories), to marketplace, to input store, and to village, hamlet, town, and all other settlements were calculated using ArcGIS software (see Fig. 2).

### 2.2.2. Field surveys

Environmental variables recorded were topographical (IV shape), related to soils (major soil types), related to hydrological functioning (water flow, emerging water table and shallow water table, and IV drainage). Major soil types and IV shapes were obtained by direct field observations or from key informants who use the IV or live close by (Fig. 2). Socio-economic candidate predictors concerned IV users, land exploitation, access to road, market, village, seeds, other inputs, extension services, and land, and farmer organizations. For farm management practices, crops, cropping system, and land development system were recorded. All socio-economic and agricultural practice variables were collected via questionnaires and observations.

### 2.2.3. Spatial candidate predictors

The spatial predictive variables (refer to Fig. 2 and Appendix) concerned flow accumulation (hydrological data), elevation (topographical data), rainfall (one component of climate), soil physical and chemical properties (organic carbon content, total nitrogen, exchangeable bases, pH in $H_2O$, sand and clay fractions), and population density (demographic data). All these were derived from online and

**Table 1**
Sampling of study sites (inland valleys, IVs) and criteria.

| | Study area | | | |
|---|---|---|---|---|
| | Mono and Couffo departments, Benin | Upper Ouémé catchment, Benin | Sikasso and Kadiolo circles, Mali | Bo and Kenema districts, Sierra Leone |
| IVs with agricultural use: | 98 | 145 | 99 | 129 |
| Paddy cultivation | 43 | 138 | 86 | 114 |
| Vegetable crops | 86 | 68 | 53 | 82 |
| Other crops | 86 | 66 | 41 | 46 |
| IVs without agricultural use | 2 | 4 | 1 | 21 |
| Total IVs surveyed (*n*) | 100 | 149 | 100 | 150 |

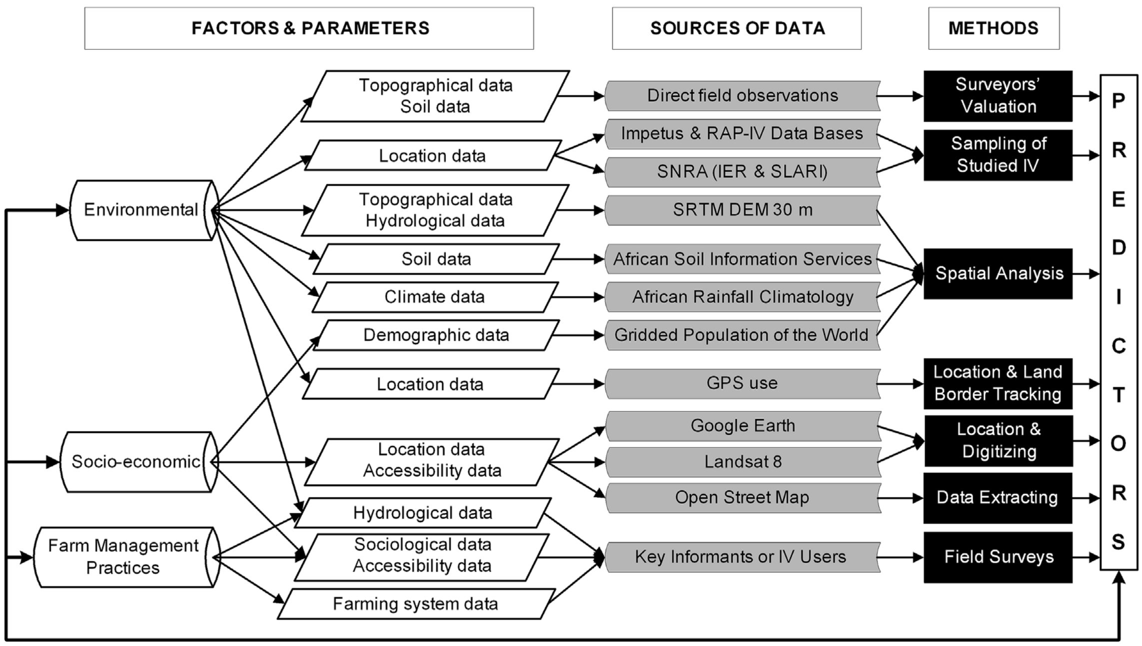


**Fig. 2.** Flow chart depicting collection and deriving of data flow.

freely available data sources using spatial analysis in ArcGIS and extracted using the geographic location of the IVs.

Worldwide High-resolution Shuttle Radar Topography Mission (SRTM 30 m) was used to establish the hydrological and topographical data sets. The maximum flow accumulation and the mean elevation in and around each IV were recorded. Daily rainfall data covering the period 1983–2015 from African Rainfall Climatology (ARC) Version 2 for Famine Early Warning Systems were used to prepare the rainfall data (Dembélé & Zwart, 2016; Novella & Thiaw, 2013). The annual average rainfall was aggregated by calculating annual sums of daily rainfall grids. Soil properties were obtained from AfSoilGrids250 m of Africa Soil Information Services (AfSIS) were aggregated (Hengl et al., 2015). Using the composite of grids for 0–5 cm, 5–15 cm, and 15–30 cm depths, means of pixel values for depth 0–30 cm for soil pH, fractions of sand and clay, total exchangeable bases (Ca, K, Mg, Na), soil organic carbon, and the total nitrogen for a depth of 0–20 cm were determined.

The Gridded Population of the World (GPW) Version 4 of 2015 with an output resolution of 30 arc-seconds provided the population density (persons/km$^2$). This was considered the maximum population density per administrative unit in which each IV is located – '*arrondissement*,' 'commune,' and 'chiefdom' in Benin, Mali, and Sierra Leone, respectively.

### 2.3. Methodological approach and data processing

#### 2.3.1. Random Forests

The statistical method Random Forests is an ensemble learning technique that builds multiple 'trees' based on random bootstrapped samples of the training data (Breiman, 2001). Having the ability to identify informative variables (Hapfelmeier & Ulm, 2013), Random Forests shows excellent performance compared to other classification methods (Zhou, Hong, Luo, & Yang, 2010). It is known as a variable selection method based on the algorithmic approach (Sandri & Zuccolotto, 2006), Random Forests can be applied when many potential predictors exist, and has good predictive performance (Tang et al., 2009). Random Forests provides measures for each variable's predictive importance (Yang & Gu, 2009).

In medicine, extensions of this method have been proposed, and aimed at identifying variables important to the trait of interest (Barco et al., 2012; Boulesteix, Janitza, Kruppa, & König, 2012; Casanova et al., 2014; Chang & Yang, 2013; Chen & Ishwaran, 2012; Chen & Wu, 2012; Díaz-Uriarte & Alvarez de Andrés, 2006; Tang et al., 2009; Zhou et al., 2010). In ecology, Random Forests has been used to deliver some significant results through satellite images using remote sensing and GIS techniques (Cutler et al., 2007; Gislason et al., 2006; Li, Tran, & Siwabessy, 2016; Loosvelt et al., 2012; Mellor, Haywood, Stone, & Jones, 2013). It has also been extensively used in agriculture (Hengl et al., 2015; Jeong et al., 2016; Laborte et al., 2012; Ozdarici-Ok, Ozgun Ok, & Schindler, 2015; Vintrou et al., 2012; Watts & Lawrence, 2008).

The objectives of variable selection are: (1) to find important variables strongly related to the response variable for interpretation purposes; and (2) to find a small number of variables sufficient for a good prediction of the response variable (Genuer, Poggi, & Tuleau-Malot, 2010). The performance of Random Forests to select important variables (Gregorutti, Michel, & saint-Pierre, 2017) has been tested in comparison with other classical means of variable selection (Genuer et al., 2010; Ozdarici-Ok et al., 2015; Sandri & Zuccolotto, 2006; Tang et al., 2009; Yang & Gu, 2009). Each variable in its original data set could be evaluated many times within different groups of variables. Globally, important variables could be selected after many repetitions (Yang & Gu, 2009). Consequently, Random Forests procedure returns a small set of predictors which have high importance and jointly give good prediction rate.

#### 2.3.2. Variable importance measures and selection of predictors

The selection of important variables is based on recursive elimination of variables (Hapfelmeier & Ulm, 2013). The importance of each variable is calculated by taking the difference between the prediction accuracy with and without permuting the variable, and then averaging this difference over all trees and normalizing by the standard error. The variables are ranked on the basis of this importance measure, with the variables having the highest decrease in accuracy resulting from the permutation identified as the most important. Variables with the lowest ranking (i.e. least important predictors) at a certain threshold are subsequently removed from the predictor set and Random Forests retrained (Laborte et al., 2012). Selection results are based on the scaled Mean Decrease in Accuracy (MDA) measure of variable importance (Calle & Urrea, 2010; Cutler et al., 2007; Jeong et al., 2016; Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012; Stevens, Gaughan, Linard, & Tatem, 2015).

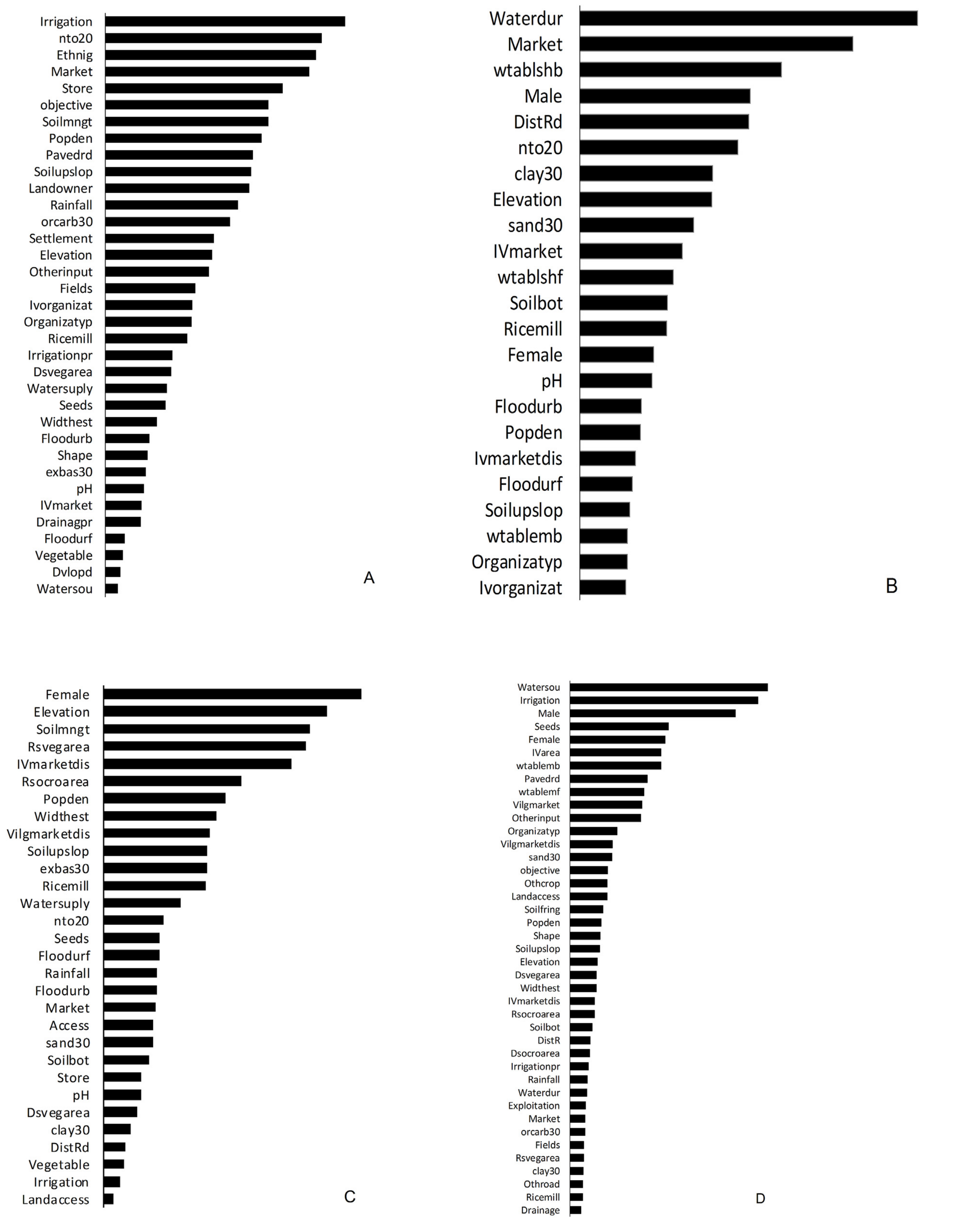


**Fig. 3.** Predictors' importance measures based on Mean Decrease in Accuracy (MDA). A: Mono and Couffo *départements*, Benin; B: Upper Ouémé catchment, Benin; C: Sikasso and Kadiolo circles, Mali and; D: Bo and Kenema districts, Sierra Leone.

The training followed several steps as advised in the recursive feature elimination (RFE) algorithm for variable selection, which has been shown to be efficient for selecting a small number of variables that together have a good prediction error (Gregorutti et al., 2017). At the start (step 1), the classification process was repeated many times. Data training was computed repeated 50 times (Genuer et al., 2010; Vintrou et al., 2012). Average MDAs were calculated per variable. All variables with MDA values of 0.05 or less were removed (Hapfelmeier & Ulm, 2013). In step 2, the reduced data set of important variables with MDA values superior of 0.05 was subjected to a repeat of the process of the first step. The process was repeated many times until no variables with MDA $\leq 0.05$ were obtained. The short list of important variables was therefore selected and the new dataset contained only the selected important variables. Among these 'important variables' to come out of step 2, several could be highly auto-correlated. The final step (step 3) therefore tested the correlation and Principal Components Analysis (PCA) of these important variables to remove the correlated variables from the list (Genuer et al., 2010; Gregorutti et al., 2017; Millard & Richardson, 2015; Nicodemus, Malley, Strobl, & Ziegler, 2010; Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008). The significant variables in the final short list were ranked as very important, moderately important, and of little importance by considering last values of MDA.

#### 2.3.3. Prediction performance evaluation

Random Forests returns a measure of error rate based on the out-of-bag cases for each fitted tree, the OOB error. The OOB error was used to evaluate the effect of changes in parameters of Random Forests on its classification of variables (Díaz-Uriarte & Alvarez de Andrés, 2006). OOB error was used (1) to compare Random Forests outputs instead of assessing models, (2) to give fair estimation compared to the usual alternative test set error it is considered somewhat optimistic, and (3) because it is a default output of Random Forests technique (Genuer et al., 2010; Millard & Richardson, 2015). The mean values of OOB error per study region were subjected to statistical test by submitting OOB error of training series to Two-way analysis of variance (ANOVA).

The Random Forests model training considers a part of the sample ($n$) which represents the predicted value. This predicted value can vary by increasing (which is an improvement of predicted value rate) of a training series to another one according to data subset or may be stable whatever the size or the variation of data subsets.

## 3. Results

### 3.1. Variable importance measures

#### 3.1.1. Variable classification and relevant predictors

The clustered bar charts presented in Fig. 3 show the classification per region of parameters measured according to their importance. Fig. 4 presented the series of 50 training per region before obtaining these classifications exhibited by clustered bar charts; five series of runs for Mono and Couffo *départements* in Benin, Upper Ouémé catchment in Benin, and Sikasso and Kadiolo *cercles* in Mali and four series of runs for Bo and Kenema districts in Sierra Leone. This variation in numbers of training runs was due to the threshold of the value of MDA used to select important variables. This criterion of selection also explains the variation in number of important variables per region. Except for Bo and Kenema districts in Sierra Leone where the first series of training favored the removal of fewer candidate predictors (3% of the initial data set), some 35%, 39%, and 47% of variables for Mono and Couffo *départements*, Upper Ouémé catchment, and Sikasso and Kadiolo *cercles*, respectively, were removed from the initial data sets. Subsequent training series on the Upper Ouémé catchment data set resulted in the removal of more variables and the candidate important predictors for this region were around a third (36%) of the initial data set, fewer than in other cases. Some 47%, 58%, and 93% of the initial data set were identified as relevant variables for Sikasso and Kadiolo *cercles*, Mono and Couffo *départements*, and Bo and Kenema districts, respectively, and were submitted to correlated test before final selection of predictors (Fig. 4).

#### 3.1.2. Prediction performance: quality of fit and predicted values

The internal model performance evaluation was focused on two model outputs: the quality of fit expressed by 'out-of-bag estimate error' rates (OOB error) and the predicted values of observations (samples). The OOB errors used to identify the optimal subset of predictive parameters per study area and per training series are presented with their standard deviations and ANOVA results in Table 2. The mean values of OOB errors for Mono and Couffo *départements* of 24.4–28.1% were relatively higher than those of other regions, while the training series of Bo and Kenema districts, Sierra Leone present the lowest OOB error mean values (4.11% without any variation). For Upper Ouémé catchment, the mean values of OOB error ranged between 5.83 and 7.63% and those for Sikasso and Kadiolo *cercles* were 11.4–13.1%. The ANOVA showed that means of OOB error of Mono and Couffo *départements* and those of Upper Ouémé catchment and Sikasso and Kadiolo *cercles* were significantly different, while the means of Bo and Kenema districts were not different. However, the OOB error of training series per region in these three study regions (Mono and Couffo *départements,* Upper Ouémé catchment and Sikasso and Kadiolo *cercles*) were not always significantly different.

Percentages of predicted values per training series and per study region are shown on Fig. 5. Note that the departure situation of observations' size ($n$) does not undergo the change. Indeed, the predicted values from training series of subsets of different study regions were close to 70% or 100% from the first training series. For the Mono and Couffo *départements* subset, percentages of predicted values were

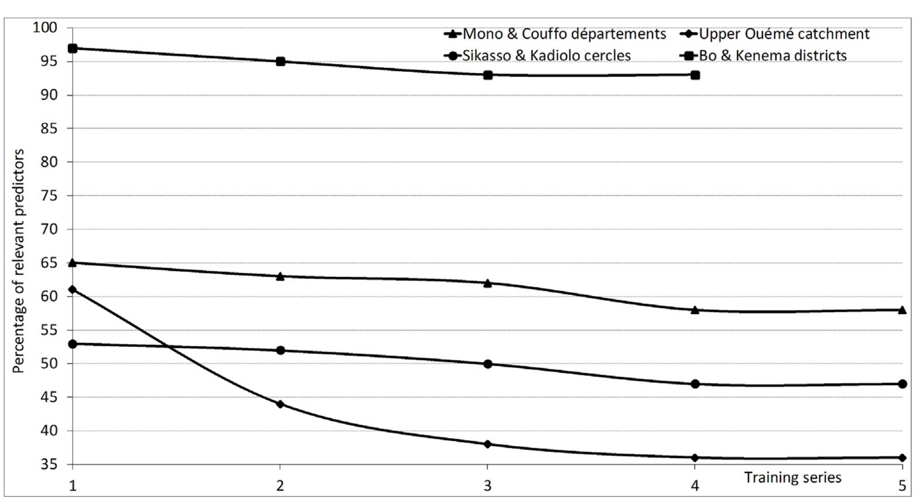


**Fig. 4.** Number of relevant predictors per training series.

**Table 2**
Out-of-bag (OOB) estimates error rates (%).

| Study area | Training series (S) | | | | | F | Bonferroni test |
|---|---|---|---|---|---|---|---|
| | S1 | S2 | S3 | S4 | S5 | | |
| Mono and Couffo départements | 28.1 (2.6) | 25.4 (2.9) | 24.4 (2.2) | 24.5 (2.1) | 24.7 (2.6) | 19.53*** | S1#S2; S1#S3; S1#S4; S1#S5; S2 = S3; S2 = S4; S2 = S5; S3 = S4; S3 = S5; S4 = S5 |
| Upper Ouémé catchment | 5.83 (0) | 7.63 (0) | 7.50 (0) | 7.50 (0) | 7.57 (0.2) | 2860.22*** | S1#S2; S1#S3; S1#S4; S1#S5; S2#S3; S2#S4; S2#S5; S3 = S4; S3#S5; S4#S5 |
| Sikasso and Kadiolo cercles | 13.1 (0) | 12.1 (0.7) | 11.9 (0.6) | 12.1 (0.6) | 11.4 (0.9) | 49.25*** | S1#S2; S1#S3; S1#S4; S1#S5; S2 = S3; S2 = S4; S2#S5; S3 = S4; S3#S5; S4#S5 |
| Bo and Kenema districts | 4.11 (0) | 4.11 (0) | 4.11 (0) | 4.11 (0) | | – | – |

() indicate Standard Deviation values.
*** indicate significant differences at $p \leq 0.05$ between OOB error means of training series.

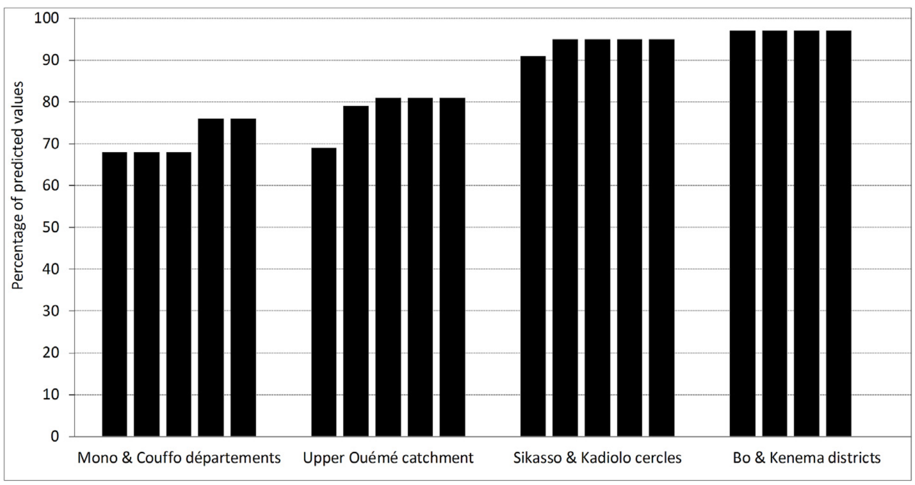


**Fig. 5.** Percentage of predicted values of training series per study areas.

between 68% and 76%, and the Upper Ouémé catchment subsets presented values between 69% and 81%. The percentages for Sikasso and Kadiolo *cercles* changed very little (91–95%) and only for first training series, while those of Bo and Kenema districts subsets were not improved at all (97%). The percentages for Bo and Kenema districts being highest and constant are in accordance with the OOB errors from that region's training series, which were lowest and constant.

### 3.2. Selection of inland valley rice production predictors

Variables determined as important by Random Forests could be categorized in three classes – 'highly,' 'moderately,' and 'weakly' important. Table 3 gives all the variables considered highly important. Socio-economic and farm management practices factors were numerically more relevant than physical parameters (soil, hydrology, and topography) for predicting the potential of IVs for rice production development. Socio-economic data essentially related to accessibility, such as the distance from the IV to the nearest market or road. The population density close to the IV, the gender (sex, ethnic group) of IV users, and sometimes availability of inputs (seeds, fertilizers) were also useful in prediction. Except for the Upper Ouémé catchment study region, farm management practices such as irrigation water resources, soil fertility management, area of vegetable crops, and type of farmer organization were also highly relevant for predicting production potential. As for physical parameters, only total soil nitrogen was highly important in the two study regions in Benin, and the soil sand fraction was highly important in Bo and Kenema districts. The water duration in IV in Upper Ouémé catchment, the water flow source in Bo and Kenema

**Table 3**
Predictors: highly important variables by study area[a].

| Factors | Mono and Couffo *départements* | Upper Ouémé catchment | Sikasso and Kadiolo *cercles* | Bo and Kenema districts |
|---|---|---|---|---|
| Socio-economic, including accessibility | Popden, Ethnig, Market, Pavedrd, Store | Male, Market, DistRd | Popden, Female, Ivmarketdis | Female, Male Seeds, Otherinput Pavedrd, Vilgmarket, Vilgmarketdis |
| Farm management practices | Irrigation, Soilmngt, objective | | Soilmngt, Rsvegarea, Rsocroarea | IVarea, Irrigation, Organizatyp |
| Soil | nto20 | nto20 | | sand30 |
| Hydrological | | Waterdur, wtablshb | | Watersou, wtablemb, wtablemf |
| Topographical | | | Elevation | |

[a] The description of variables is clearly displayed in the Appendix of this manuscript just after References.

districts, and the mean elevation of IVs in Sikasso and Kadiolo *cercles* were likewise highly important. Beyond these predictors, correlated highly important variables were also identified. These were the distance from the IV to the nearest paved road (Pavedrd) and the distance from the IV to the nearest store for inputs (Store) in Mono and Couffo *départements*; wet-season crops (other than paddy and vegetable fields) cultivation area (Rsocroarea) in Sikasso and Kadiolo *cercles*; and the number of male farmers using the IV (Male), source of chemical inputs (Otherinput), duration of an emergent water table in the IV bottom (wtablemb), and duration of emergent water table on the IV fringe (wtablemf) in Bo and Kenema districts.

The ranking criteria used were the variable's importance level (high, moderate, weak) and the representativeness of each important variable in the four study regions (either important in all study regions or important in three study regions). Table 4 presents the most relevant and representative predictors from this study. Ranking of predictors took into account the removal of correlated variables, while retaining the more important and non-correlated variables. Thus, the distance from the IV to the nearest market (Market) and the population density (Popden) were revealed as the best predictors (first order) whatever the geographical location of the IV, despite the diverse socio-economic and bio-physical conditions (Table 4). The second order comprised total soil nitrogen (nto20) and the mean elevation of the IV (Elevation). The third order comprised the number of female farmers working in the IV (Female) and the soil texture (clayey, sandy, or intermediate) upper slope of the IV (Soilupslop). The fourth-order predictors were the source (in village or 25 km, or so far) of seeds (Seeds), soil fertility management – whether or not fertilizer was used (Soilmngt), the irrigation water resource (Irrigation) which are either river diversion or natural spring or mixed, and the soil sand fraction (sand30). Distance from the IV to the nearest road (DisRd), from the IV to the nearest market (Ivmarketdis), type of farmer organization (Organizatyp), flooding duration in the IV bottom (Floodurb), soil pH, soil texture in the IV bottom (Soilbot), and estimated average width of the IV (Widthest) were relevant and representative at fifth order.

Variables such as the distance from the IV to the nearest rice mill (Ricemill) and the estimated annual average rainfall, which were also important and representative but are correlated with another variable, were removed from the list of predictors. Moreover, in each study region, predictors were revealed highly important and without any correlation with another variable, but not representative across all study areas. These location-specific predictors included the production objective (objective) and the major ethnic groups (Ethnig) in Mono and Couffo *départements*; the number of male farmers working in the IV (Male), water flow duration in the IV (Waterdur), and the duration of shallow water table in the IV bottom (wtablshb) in Upper Ouémé catchment; the wet-season vegetable cultivation area (Rsvegarea) in Sikasso and Kadiolo *cercles*; and the distance from the IV to the nearest paved road (Pavedrd), the village to market distance (Vilgmarketdis), the state of the road between the village and the market (Vilgmarket), the water flow source (Watersou), and the total area of the IV (IVarea) in Bo and Kenema districts.

## 4. Discussion and conclusions

### 4.1. Random Forests for selection of inland valley rice production predictors

According to Genuer et al. (2010), one advantage of Random Forests is that it performs well for both classic problems (where $n > p$), the case of this study, and for problems of high dimension (where $n < p$), where n is the size of the sample and p the number of variables. For all variable importance measures in Random Forests processing, two indices – MDA and mean decrease Gini (MDG) – are available to be used to classify variables. The choice depends on the study area and the objectives of the study. Most studies that apply Random Forests methods use MDA for measurement of the importance of variables. In

**Table 4**
Relevant and representative predictors per factor.

| Factors | Predictors | Importance[a] | | | | | Representativeness[b] | Ranking |
|---|---|---|---|---|---|---|---|---|
| | | Mono and Couffo *départements* | Upper Ouémé catchment | Sikasso and Kadiolo *cercles* | Bo and Kenema district | Index | | |
| Accessibility | Distance from IV to nearest market (Market) | 3 | 3 | 2 | 2 | 10 | α | 1st |
| | Distance from IV to road (DistRd) | 0 | 3 | 1 | 2 | 6 | β | 5th |
| | IV–market distance (Ivmarketdis) | 0 | 1 | 3 | 2[c] | 6 | β | 5th |
| Other socio-economic | Population density (Popden) | 3 | 2 | 3 | 2 | 10 | α | 1st |
| | Number of female farmers in the IV (Female) | 0 | 2[c] | 3 | 3 | 8 | β | 3th |
| | Source of seeds (Seeds) | 2 | 0 | 2 | 3 | 7 | β | 4th |
| Farm management practices | Soil fertility management (Soilmngt) | 3 | 0 | 3 | 1 | 7 | β | 4th |
| | Irrigation water resource (Irrigation) | 3 | 0 | 1 | 3 | 7 | β | 4th |
| | type of organization (Organizatyp) | 2 | 1 | 0 | 3 | 6 | β | 5th |
| Hydrological | Flooding duration in IV bottom (Floodurb) | 2 | 2 | 2 | 0 | 6 | β | 5th |
| Soil | Total nitrogen in top 20 cm (nto20) | 3 | 3 | 2 | 1[c] | 9 | α | 2nd |
| | Soil texture on the upper slope (Soilupslop) | 2 | 1 | 2 | 2 | 7 | α | 3th |
| | Sand content in top 30 cm (sand30) | 0 | 2 | 2 | 3 | 7 | β | 4th |
| | Soil pH in H2O (pH) | 1[c] | 2 | 1 | 1 | 5 | α | 5th |
| | Soil texture in the IV bottom (Soilbot) | 0 | 2 | 2 | 2[c] | 6 | β | 5th |
| Topo-graphical | Mean elevation (Elevation) | 2 | 2 | 3 | 2 | 9 | α | 2nd |
| | Estimated average width (Widthest) | 2 | 0 | 2 | 2 | 6 | β | 5th |

[a] 3 = highly important variables, 2 = moderately important, 1 = weakly important, and 0 = not important variables.
[b] α means the variable is important for all four studied regions (whether highly, moderately, or weakly) and β means the variable is important for at least three of the four studied regions.
[c] Variable correlated with another one.

bioinformatics, it was concluded that variable importance rankings based on MDG show sensitivity to within-predictor correlation and are more robust to small perturbations of the data (Boulesteix, Bender, Bermejo, & Strobl, 2011; Calle & Urrea, 2010; Nicodemus, 2011). The literature indicates that MDG is more often applied in medicine and bioinformatics. Many studies have revealed that MDG is affected by bias (Strobl, Boulesteix, Zeileis, & Hothorn, 2007; Sandri & Zuccolotto, 2008, 2010). For this reason, MDA index was chosen to measure variable importance by eliminating less relevant variables from the data set of each study region in this study.

About the technique of recursive feature elimination of variables, Díaz-Uriarte and Alvarez de Andrés (2006) eliminated, at each step, the 20% of predictors with the lowest MDA values and build a new 'forest' with the remaining variables. They finally selected the subset of predictors leading to the smallest OOB error. This technique to eliminate the 20% least relevant variables did not seem the most statistically sound method. However, the elimination of unimportant variables using the threshold average MDA value of 0.05 applied was not always efficient. This was the case, for example, for the database from Bo and Kenema districts in Sierra Leone. Means of OOB errors from the training series were not statistically different: MDA averages of training series for this region varied little. This is a major limitation of Random Forests selection by defining the threshold between predictive and non-predictive variables (Sandri & Zuccolotto, 2006). Despite a good classification of predictors of IV rice production development for this region, the method did not eliminate variables to distinguish the most relevant predictors. That could be due to many reasons, such as the natural or socio-economic environment of the IVs not enabling to obtain data for good modelling. A different approach might be more efficient for removing less-important variables after each training series until obtaining the most important, especially in the case of Bo and Kenema districts in Sierra Leone IV rice production development predictors.

### 4.2. Selected predictors for mapping inland valley rice production potential

The suitability of an area for rice production development in SSA can depend on many factors and parameters. It should be viewed from a deeper perspective taking account of technology, policy, and socio-economic factors. The IV agro-ecosystems considered here also have their own specificities related to the physical environment (Gumma et al., 2009; Laborte et al., 2012; Masoud et al., 2013; Nwanze, Mohapatra, Kormawa, Keya, & Bruce-Oliver, 2006). For predicting the potential for development of rice production in an IV agro-ecosystem efficiently, consideration must be given to the socio-economic, bio-physical, and farm management conditions within a natural, economic, agro-ecological, or political region. More than 60 variables (Appendix) which were assessed according to their importance for rice production in a natural West African environment were considered.

The distance from the IV to the nearest market, nearest road in general, and nearest paved road in particular, and the distance from village to market are all accessibility parameters that were revealed as highly important variables and therefore main predictors (Table 3). Certain distances (Market and Pavedrd) were estimated from spatial analysis. Others (DistRd, Ivmarketdis, and Vilgmarketdis) were estimated on the basis of farmers' knowledge. These parameters of distance although estimated differently are highly important for predicting the suitability for rice production development of an IV. Distances complement one another and the parameter 'distance' from rice cultivation area to market and to roads would be more relevant predictors close to the population density around the IV, total soil nitrogen, and land elevation of the IV, which are also highly relevant predictors. Studies have shown that access to roads favors agricultural intensification in IVs (Erenstein, 2006; Erenstein, Oswald, & Mahaman, 2006). In rural environments with good road access, farmers plant fewer crops, purchase more fertilizer, and hire more labor (Qin & Zhang, 2016). In the case of IVs used for agriculture and even rice production, farmers easily reach markets and stores for buying agricultural inputs (seeds and chemical fertilizers) and selling agricultural products; they can also access agricultural technologies such as rice mills, improved varieties, irrigation or water control techniques, soil management knowledge, farmer organizations, and can take advantage of agricultural extension services. Therefore, road connections improve household agricultural income, reduce poverty, and significantly increase local non-farm income for poor households (Qin & Zhang, 2016; Sharma, 2016). This explains the impact of the population density around or close to an IV. Access to labor, technologies, and markets for selling farm products such as rice (local consumption) depend on the size and the quality of the local population, which may limit or be a great advantage for rice production development. There would also be a relationship between local population density and the crops cultivated as potential requirements of the neighboring population. Thus, the increasing of the population density is important factor in agricultural development (Meertens, Fresco, & Stoop, 1996).

Hydrological factors proved important in certain cases. For example, the duration of water flow in the IV and the duration of the shallow water table at the IV bottom in Upper Ouémé catchment, and the water flow source and durations of emerging water table at the bottom and on the fringes of IVs in Bo and Kenema districts. However, hydrological factors were not strong determinants of the development of rice production in all IVs. IVs in the two regions concerned are generally small: average 6 ha and 19 ha in Bo and Kenema districts and Upper Ouémé catchment, respectively, compared with 46 ha in Mono and Couffo *départements* southern Benin and more than 100 ha in Sikasso and Kadiolo *cercles*.

In considering all natural and socio-economic factors, this study was able to identify the most relevant predictors for rice cultivation development of IVs in many agro-ecological zones of four West African regions. However, the similarities between the four regions were few. Except for accessibility parameters, it did not find the same variable highly relevant as predictor in more than two of the four studied regions. Thus, it appears necessary to combine many criteria for ranking the predictors in the study area. However, the importance of predictors (highly, moderately, and weakly) simultaneously for the four studied regions was considered. The index derived provided the first level of ranking, which is the most important. The representativeness and ranking of a predictor across regions was determined on its important in all four regions or at least in three regions. From this scenario, 13 classes of predictors were obtained, of which the best classes are presented in Table 4. Among all these predictors, the best for predicting the potential for rice production in IV agro-ecosystems of the study area were: distance from the IV to the nearest market (Market), population density (Popden), mean elevation of the IV (Elevation), and total soil nitrogen (nto20). Secondarily, other six predictors: upper-slope soil texture (Soilupslop), number of female farmers working in the IV (Female), soil fertility management (Soilmngt), irrigation water resource (Irrigation), source of seeds (Seeds), and soil sand fraction (sand30), which can also be used to effectively predict the potential for rice production development in the study area despite not being the best predictors were considered. The distance from the IV to the nearest rice mill (Ricemill) was identified as a good predictor, because it was a moderately important variable in all four regions, but was discarded because it was correlated with other (highly important) variables in each of the four regions. All other variables, despite their importance revealed by Random Forests model training shown in Table 3, would predict rice development for IVs only in the regions for which they were highly important.

## Data availability

The data base that was developed and deployed in this study is made available to the research community in a separate publication in the journal *Data in Brief* (Djagba, Kouyaté, Baggie, & Zwart, 2018a).

**Acknowledgements**

This study was supported by the European Commission through the International Fund for Agricultural Development [grant number C-ECG-65-WARDA]; and the Global Rice Science Partnership (GRiSP).

The authors are grateful to the many agents of Sierra Leone Agricultural Research Institute, Institut d'économie rurale of Mali, and Ministry of Agriculture of Benin for conducing field data collection. Special thanks to AfricaRice Biometricians, Dr. Ibnou Dieng and Amakoe Delali Alognon, for conceptualizing the graphical interface of R Software (ARiS for Windows) containing randomForest package used as our main statistical tool for this paper. The authors would also like to thank Elliott Dossou-Yovo, Guy Manners and Wilfried Yergo for their helpful supports on the manuscript.

## Appendix. Variables used as model inputs – candidate predictors

| Factors & Parameters | Variables | Description | Unit | Type | Source |
|---|---|---|---|---|---|
| Hydrological data | Floodurf | Flooding duration in inland valley (IV) fringe | Week | Quantitative | Field survey |
| | Floodurb | Flooding duration in IV bottom | Week | Quantitative | Field survey |
| | Flowacc | Flow accumulation (maximum) | Index | Quantitative | DEM/STRM[a] (30 m) |
| | Watersou | Water flow source | | Qualitative | Field survey |
| | Waterdur | Water flow duration | | Qualitative | Field survey |
| | Watflodur | Water flow duration if temporary | Month | Quantitative | Field survey |
| | wtablemb | Emerging water table, IV bottom duration | Month | Quantitative | Field survey |
| | wtablemf | Emerging water table, IV fringe duration | Month | Quantitative | Field survey |
| | Wtablshb | Shallow water table, IV bottom duration | Month | Quantitative | Field survey |
| | Wtablshf | Shallow water table, IV fringe duration | Month | Quantitative | Field survey |
| | Drainage | IV drainage | | Qualitative | Field survey |
| Topographical and climatic data | Shape | Transversal entrenchment shape | | Qualitative | Field survey |
| | Elevation | Elevation (mean) | Meter | Quantitative | DEM/STRM (30 m) |
| | Widthest | Estimated average width | Meter | Quantitative | Field survey |
| | Rainfall | Estimated annual average rainfall | Millimeter | Quantitative | ARC2 for FEWS[b] |
| Soil data | OC | Soil organic carbon content | $g\ kg^{-1}$ | Quantitative | AfSoilGrids250m[c] |
| | nto20 | Total nitrogen in top 20 cm | $g\ kg^{-1}$ | Quantitative | AfSoilGrids250 m |
| | Exchbas | Exchangeable bases in top 30 cm | Cmolc $kg^{-1}$ | Quantitative | AfSoilGrids250 m |
| | sand30 | Sand content in top 30 cm | Percent | Quantitative | AfSoilGrids250 m |
| | clay30 | Clay content in top 30 cm | Percent | Quantitative | AfSoilGrids250 m |
| | pH | Soil pH in H2O | Index | Quantitative | AfSoilGrids250 m |
| | Soilbot | Soil texture in the IV bottom | | Qualitative | Field survey |
| | Soilfring | Soil texture on the IV fringe | | Qualitative | Field survey |
| | Soilupslop | Soil texture on the upper slope | | Qualitative | Field survey |
| Socio-economic and accessibility | Pavedrd | Distance from IV to nearest paved road | Meter | Quantitative | OSM[d] & GoogleEarth |
| | Othroad | Distance from IV to nearest other road | Meter | Quantitative | OSM & GoogleEarth |
| | DistRd | Distance from IV to road | km | Quantitative | Field survey |
| | Settlement | Distance from IV to nearest settlement | Meter | Quantitative | OSM & GoogleEarth |
| | Market | Distance from IV to nearest market | Meter | Quantitative | GPS location |
| | Ricemill | Distance from IV to nearest rice mill | Meter | Quantitative | GPS location |
| | Store | Distance from IV to nearest store of inputs | Meter | Quantitative | GPS location |
| | IVmarket | Road IV–market | | Qualitative | Field survey |
| | Vilgmarket | Road village–market | | Qualitative | Field survey |
| | IVmarketdis | IV–market distance | km | Quantitative | Field survey |
| | Vilgmarketdis | Village–market distance | km | Quantitative | Field survey |
| | Popden | Population density | Person $km^{-2}$ | Quantitative | GPWV4[e] |
| | Landowner | Land ownership | | Qualitative | Field survey |
| | Male | Number of male farmers in the IV | Person | Quantitative | Field survey |
| | Female | Number of female farmers in the IV | Person | Quantitative | Field survey |
| | Ethnig | Major ethnic groups | | Qualitative | Field survey |
| | Migranpred | Predominance of migrants in use of IV | | Qualitative | Field survey |
| | Landaccess | Access to land | | Qualitative | Field survey |
| | Access | Accessibility of the IV | | Qualitative | Field survey |
| | Seeds | Source of seeds | | Qualitative | Field survey |
| | Otherinput | Source of other inputs | | Qualitative | Field survey |

| | | | | | |
|---|---|---|---|---|---|
| Farm management practices data | Othcrop | Other crops in IV | | Qualitative | Field survey |
| | Vegetable | Vegetables in IV | | Qualitative | Field survey |
| | IVarea | Total area of the IV | Hectare | Quantitative | GPS data/ GoogleEarth |
| | Exploitation | Mode of exploitation | | Qualitative | Field survey |
| | Objective | Production objective | | Qualitative | Field survey |
| | Agrisupport | Presence of agricultural support structure | | Qualitative | Field survey |
| | Ivorganizat | Existence of IV farmer organization | | Qualitative | Field survey |
| | Organizatyp | If yes, type of organization and if no, none | | Qualitative | Field survey |
| | Dvlopd | Is IV developed? | | Qualitative | Field survey |
| | Soilmngt | Soil fertility management | | Qualitative | Field survey |
| | Watersuply | Water supply | | Qualitative | Field survey |
| | Irrigation | Irrigation water resource | | Qualitative | Field survey |
| | Fields | Field development | | Qualitative | Field survey |
| | Drainagpr | Drainage practices | | Qualitative | Field survey |
| | Irrigationpr | Irrigation practices | | Qualitative | Field survey |
| | Rsvegarea | Wet-season vegetable cultivation area | Hectare | Quantitative | Field survey |
| | Dsvegarea | Dry-season vegetable cultivation area | Hectare | Quantitative | Field survey |
| | Rsocroarea | Wet-season other crops cultivation area | Hectare | Quantitative | Field survey |
| | Dsocroarea | Dry-season other crops cultivation area | Hectare | Quantitative | Field survey |

[a] Digital Elevation Model/Worldwide High-resolution Shuttle Radar Topography Mission (SRTM 30 m), URL: http://srtm.csi.org Data derivation were done in ArcGIS.

[b] African Rainfall Climatology Version 2 for Famine Early Warning Systems available at ftp.cpc.ncep.noaa.gov/fews/fewsdata/africa/arc2.

[c] Soil properties of African at 250 m, Soil Grids available at www.isric.org/data/AfSoilGrids250m.

[d] Open Street Map or digitizing from Google Earth. Layers derivation were done in ArcGIS.

[e] Gridded Population of the World (GPW) Version 4 in 2015, Center for International Earth Science Information Network (CIESIN).

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apgeog.2018.05.003.

## References


Abe, S. S., Buri, M. M., Issaka, R. N., Kiepe, P., & Wakatsuki, T. (2010). Soil fertility potential for rice production in West Africa lowlands. *Japan Agricultural Research Quarterly, 44*, 343–355.

AfricaRice (2014). *Africa rice trends 2001-2010.* Cotonou, Benin: Africa Rice Center (AfricaRice)108.

Barco, L., Mancin, M., Ruffa, M., Saccardin, C., Minorello, C., Zavagnin, P., et al. (2012). Application of the Random Forest method to analyse epidemiological and phenotypic characteristics of Salmonella 4,[5],12:i:- and Salmonella Typhimurium Strains. *Zoonoses and Public Health, 59*, 505–512.

Boulesteix, A.-L., Bender, A., Bermejo, J. L., & Strobl, C. (2011). Random forest Gini importance favours SNPs with large minor allele frequency: Impact, sources and recommendations. *Briefings in Bioinformatics, 13*, 292–304.

Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Mining and Knowledge Discovery, 2*, 493–507.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Brown, M. E., Hintermann, B., & Higgins, N. (2009). Markets, climate change, and food security in West Africa. *Environmental Science & Technology, 43*, 8016–8020.

Calle, M. L., & Urrea, V. (2010). Letter to the Editor: Stability of random forest importance measures. *Briefings in Bioinformatics, 12*, 86–89.

Casanova, R., Saldana, S., Chew, E., Danis, R., Greven, C., & Ambrosius, W. T. (2014). Application of random forests methods to diabetic retinopathy classification analyses. *PLoS One, 9*, e98587.

Chang, K., & Yang, J.-R. (2013). Analysis and prediction of highly effective antiviral peptides based on random forests. *PLoS One, 8*, e70166.

Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics, 99*, 323–329.

Chen, C.-L., & Wu, C.-W. (2012). Diagnosing assets impairment by using Random Forests. *International Journal of Information Technology and Decision Making, 11*, 77–102.

Cutler, D. R., Edwards, T. C., Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random forests for classification in ecology. *Ecology, 88*, 2783–2792.

Danvi, A., Giertz, S., Zwart, S. J., & Diekkrüger, B. (2018). Rice intensification in a changing environment: Impact on water availability in Inland valley landscapes in Benin. *Water, 10*, w74. http://dx.doi.org/10.3390/w10010074.

Danvi, A., Jütten, T., giertz, S., Zwart, S. J., & Diekkruger, B. (2016). A spatially explicit approach to assess the suitability for rice cultivation in an inland valley in central Benin. *Agricultural Water Management, 177*, 95–106.

Dembélé, M., & Zwart, S. J. (2016). Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. *International Journal of Remote Sensing, 37*, 3995–4014.

Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics, 7*, 3.

Djagba, J. F., Kouyaté, A. M., Baggie, I., & Zwart, S. J. (2018a). *A geospatial database of inland valley surveys in 4 zones in Benin, Sierra Leone and Liberia.* Data in Brief (under review).

Djagba, J. F., Rodenburg, J., Zwart, S. J., Houndagba, C. J., & Kiepe, P. (2014). Failure and success factors of irrigation system developments - case study from the Ouémé and Zou valleys in Benin. *Irrigation and Drainage, 63*, 328–339.

Djagba, J. F., Zwart, S. J., Houssou, C. S., Tenté, B. H. A., & Kiepe, P. (2018b). Ecological sustainability and environmental risks of inland valley agro-ecosystems development in Benin. *Environment, Development and Sustainability*. http://dx.doi.org/10.1007/s10668-10018-10107 (in press).

Dossou-Yovo, E. R., Baggie, I., Djagba, J. F., & zwart, S. J. (2017). Diversity of inland valleys and opportunities for agricultural development in Sierra Leone. *PLoS One, 12*(6) e0180059.

Erenstein, O. (2006). Intensification or extensification? Factors affecting technology use in peri-urban lowlands along an agro-ecological gradient in West Africa. *Agricultural Systems, 90*, 132–158.

Erenstein, O., Oswald, A., & Mahaman, M. (2006). Determinants of lowland use close to urban markets along an agro-ecological gradient in West Africa. *Agriculture, Ecosystems & Environment, 117*, 205–217.

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*, 2225–2236.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters, 27*, 294–300.

Gregorutti, B., Michel, B., & saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing, 27*, 659–678.

Gruber, I., Kloss, J., & Shopp, M. (2009). Seasonal water demand in Benin's agriculture. *Journal of Environmental Management, 90*, 196–205.

Gumma, M., Thenkabail, P. S., Fujii, H., & Namara, R. (2009). Spatial models for selecting the most suitable areas of rice cultivation in the Inland Valley Wetlands of Ghana using remote sensing and geographic information systems. *Journal of Applied Remote Sensing, 3*, 1–21.

Hapfelmeier, A., & Ulm, K. (2013). A new variable selection approach using Random Forests. *Computational Statistics & Data Analysis, 60*, 50–69.

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., et al. (2015). Mapping soil properties of Africa at 250 m Resolution: Random forests significantly improve current predictions. *PLoS One, 10* e0125814.

Inocencio, A., Kikuchi, M., Tonosaki, M., Maruyama, A., Merrey, D., Sally, H., et al. (2007). *Costs and performance of irrigation projects: A comparison of sub-Saharan Africa and other developing regions.* Colombo, Sri Lanka: International Water Management Institute (IWMI)81 Research Report 109.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., et al. (2016). Random forests for global and regional crop yield predictions. *PLoS One, 11* e0156571.

Laborte, A. G., Maunahan, A. A., & Hijmans, R. J. (2012). Opportunities for expanding paddy rice production in Laos: Predictive modeling using random forest. *Journal of Land Use Science, 7*, 21–33.

Li, J., Tran, M., & Siwabessy, J. (2016). Selecting optimal random forest predictive models: A case study on predicting the spatial distribution of seabed hardness. *PLoS One, 11* e0149089.

Loosvelt, L., Peters, J., Skriver, H., Lievens, H., Van Coillie, F. M. B., De Baets, B., et al. (2012). Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification. *International Journal of Applied Earth Observation and Geoinformation, 19*, 173–184.

Masoud, J., Agyare, W. A., Forkuor, G., Namara, R., & Ofori, E. (2013). Modeling inland valley suitibility for rice cultivation. *ARPN Journal of Engineering and Applied Sciences, 8*, 9–19.

McCartney, M. P., & Houghton-Carr, H. A. (2009). Working Wetland Potential: An index to guide the sustainable development of African wetlands. *Natural Resources Forum, 33*, 99–110.

Meertens, H. C. C., Fresco, L. O., & Stoop, W. A. (1996). Farming systems dynamics: Impact of increasing population density and the availability of land resources on changes in agricultural systems. The case of Sukumaland, Tanzania. *Agriculture, Ecosystems & Environment, 56*, 203–215.

Meijer, S. S., Catacutan, D., Ajayi, O. C., Sileshi, G. W., & Nieuwenhuis, M. (2015). The role of knowledge, attitudes and perceptions in the uptake of agricultural and agroforestry innovations among smallholder farmers in sub-Saharan Africa. *International Journal of Agricultural Sustainability, 13*, 40–54.

Mellor, A., Haywood, A., Stone, C., & Jones, S. (2013). The performance of Random Forests in an operational setting for large area sclerophyll forest classification. *Remote Sensing, 5*, 2838–2856.

Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing, 7*, 8489–8515.

Musa, I. K. (2009). Priority issues of least developed countries in Africa. *Irrigation and Drainage Systems, 58*, 39–51.

Niang, A., Becker, M., Ewert, F., Dieng, I., Gaiser, T., Tanaka, A., et al. (2017). Variability and determinants of yields in rice production systems of West Africa. *Field Crops Research, 207*, 1–12.

Nicodemus, K. K. (2011). Letter to the Editor:On the stability and ranking of predictors from random forest variable importance measures. *Briefings in Bioinformatics, 12*, 369–373.

Nicodemus, K. K., Malley, J. D., Strobl, C., & Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinformatics, 11*, 110.

Novella, N. S., & Thiaw, W. M. (2013). African rainfall Climatology version 2 for famine early warning systems. *Journal of Applied Meteorology and Climatology, 52*, 588–606.

Nwanze, K. F., Mohapatra, S., Kormawa, P., Keya, S., & Bruce-Oliver, S. (2006). Perspective - rice development in sub-saharan Africa. *Journal of the Science of Food and Agriculture, 86*, 675–677.

Nwite, J. C., Obalum, S. E., Igwe, C. A., & Wakatsuki, T. (2016). Soil physical properties and grain yields of lowland rice in response to sawah preparation intensities and soil amendment types. *Biological Agriculture & Horticulture, 32*, 192–205.

Obalum, S. E., Nwite, J. C., Oppong, J., Igwe, C. A., & Wakatsuki, T. (2011). Comparative topsoil characterization of sawah rice fields in selected inland valleys around bida, north-central Nigeria: Textural, structural and hydrophysical properties. *Paddy and Water Environment, 9*, 291–299.

Ozdarici-Ok, A., Ozgun Ok, A., & Schindler, K. (2015). Mapping of agricultural crops from single high-resolution multispectral images—data-driven smoothing vs. Parcel-Based smoothing. *Remote Sensing, 7*, 5611–5638.

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and Random Forests for ecological prediction. *Ecosystems, 9*, 181–199.

Qin, Y., & Zhang, X. (2016). Road to specialization in agricultural production: Evidence from rural China. *World Development, 77*, 1–16.

Rodenburg, J., Zwart, S. J., Kiepe, P., Narteh, L. T., Dogbe, W., & Wopereis, M. C. S. (2014). Sustainable rice production in African inland valleys: Seizing regional potentials through local approaches. *Agricultural Systems, 123*, 1–11.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93–104.

Sakané, N., Alvarez, M., Becker, M., Böhme, B., Handa, C., Kamiri, H. W., et al. (2011). Classification, characterization, and use of small wetlands in East Africa. *Wetlands, 31*, 1103–1116.

Sandri, M., & Zuccolotto, P. (2006). Variable selection using random forests. In S. Zani, A. Cerioli, M. Riani, & M. Vichi (Eds.). *Data analysis, classification and the forward search* (pp. 263–270). Springer Berlin Heidelberg.

Sandri, M., & Zuccolotto, P. (2008). A bias correction algorithm for the Gini variable importance measure in classification trees. *Journal of Computational & Graphical Statistics, 17*, 611–628.

Sandri, M., & Zuccolotto, P. (2010). Analysis and correction of bias in Total Decrease in Node Impurity measures for tree-based algorithms. *Statistics and Computing, 20*, 393–407.

Sasson, A. (2012). Food security for Africa: An urgent global challenge. *Agriculture & Food Security, 1*, 1–16.

Seck, P. A., Diagne, A., Mohanty, S., & Wopereis, M. C. S. (2012). Crops that feed the world 7: Rice. *Food Security, 4*, 7–24.

Seck, P. A., Tollens, E., Wopereis, M. C. S., Diagne, A., & Bamba, I. (2010). Rising trends and variability of rice prices: Threats and opportunities for sub-Saharan Africa. *Food Policy, 35*, 403–411.

Sharma, A. (2016). Urban proximity and spatial pattern of land use and development in rural India. *Journal of Development Studies*. http://dx.doi.org/10.1080/00220388.2016.1166207.

Sintondji, L. O., Huat, J., Dossou-Yovo, E. R., Fusillier, J.-L., Agbossou, E. K., Djagba, J. F., et al. (2016). Scientific research and essays lessons withdrawn from the diversity of inland valleys cultivation at a regional scale: A case study of Mono and Couffo departments in south Benin. *Scientific Research and Essays, 11*, 221–229.

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One, 10* e0107042.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*, 307.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*, 25.

Tang, R., Sinnwell, J. P., Li, J., Rider, D. N., de Andrade, M., & Biernacka, J. M. (2009). Identification of genes and haplotypes that predict rheumatoid arthritis using random forests. *BMC Proceedings, 3*(Suppl. 7), S68.

Van Oort, P. A. J., & Zwart, S. J. (2018). Impacts of climate change on rice production in Africa and causes of simulated yield changes. *Global Change Biology, 24*, 1029–1045.

Vintrou, E., Soumaré, M., Bernard, S., Bégué, A., Baron, C., & Lo Seen, D. (2012). Mapping fragmented agricultural systems in the sudano-sahelian environments of Africa using random forest and ensemble metrics of coarse resolution MODIS imagery. *Photogrammetric Engineering & Remote Sensing, 78*, 839–848.

Watts, J. D., & Lawrence, R. L. (2008). Merging Random Forest classification with an object-oriented approach for analysis of agricultural lands. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 579–582 XXXVII.

Windmeijer, P. N., & Andriesse, W. (Vol. Eds.), (1993). *Inland valleys in West Africa: An agroecological characterization of rice growing environments: Vol. 52*, (pp. 160p–). Wageningen: ILRI Publication.

Worou, O. N., Gaiser, T., Saito, K., Goldbach, H., & Ewert, F. (2012). Simulation of soil water dynamics and rice crop growth as affected by bunding and fertilizer application in inland valley systems of West Africa. *Agriculture, Ecosystems & Environment, 162*, 24–35.

Xie, H., You, L., Wielgosz, B., & Ringler, C. (2014). Estimating the potential for expanding smallholder irrigation in Sub-Saharan Africa. *Agricultural Water Management, 131*, 183–193.

Yang, W., & Gu, C. C. (2009). Selection of important variables by statistical learning in genome-wide association analysis. *BMC Proceedings, 3*(Suppl. 7), S70.

Zhou, Q., Hong, W., Luo, L., & Yang, F. (2010). Gene selection using random forest and proximity differences criterion on DNA microarray data. *Journal of Convergence Information Technology, 5*, 161–170.