Mapping explanatory drivers of persistent undernutrition in the Sahelian region using machine-learning methods

April 2024

Nutrition Research Facility
Knowledge and Research for Nutrition

About the Nutrition Research Facility

The Knowledge and Research for Nutrition project of the European Commission (2020-2026) aims to provide improved knowledge and evidence for policy and program design, management and monitoring & evaluation in order to reach better nutrition outcomes.

The project is implemented by Agrinatura - the European Alliance on Agricultural Knowledge for Development – which has established a Nutrition Research Facility, pooling expertise from European academia and having the ability to mobilize internationally renowned scientific networks and research organizations from partner countries.

The Nutrition Research Facility provides expert advice to the European Commission and to the European Union (EU) Member States and Partner Countries.

Contact: nrf@agrinatura-eu.eu

Disclaimer

This publication was produced with the financial support of the European Union. Its contents are the sole responsibility of AGRINATURA and do not necessarily reflect the views of the European Union.

To cite this report:

Oliveira, S., Morgado, P., Velhinho, H., Rita Morais, A., Capinha, C., Serras, J. (April 2024). Mapping explanatory drivers of persistent undernutrition in the Sahelian region using machine-learning methods (report). Nutrition Research Facility.

Document information

Deliverable	3			
Work Package	2			
Nature	Evidence-based final report with the identification of the main drivers			
Lead Author	Sandra Oliveira (co-Leader)			
Contribution	Hugo Velhinho (Researcher), Ana Rita Morais (Researcher), César Capinha (Researcher), João Serras (Consultant), Paulo Morgado (Leader)			
Reviewer(s)	Gabriela Albuquerque (NKE – NRF WP2 research assistant), Luís Goulão (KE – NRF WP2 leader), Carl Lachat (NRF KE - Quality Assurance), Paolo Sarfatti (KE – NRF Team Leader)			
Date of Delivery	**Contractual**	30.06.2023	**Actual**	30.06.2023

Document history

Version	**Issue date**	**Stage**	**Changes**	**Contribution**
1.0	30.06.2023			Gabriela Albuquerque Luís Goulão Paolo Sarfatti Ravinder Kumar
2.0	04.10.2023			Gabriela Albuquerque Luís Goulão Paolo Sarfatti
3.0	12.01.2024		Clarifications after NRF QA	Carl Lachat
4.0	23.04.2024		Clarification of some concepts and improvement of graphical aspects	Gabriela Albuquerque Luís Goulão Paolo Sarfatti

List of Acronyms

Acronym	Description
AI	Artificial Intelligence
BRT	Boosted Regression Trees
DHS	Demographic and Health Survey
GIS	Geographic Information System
ML	Machine Learning
PCC	Pearson Correlation Coefficient
RAE	Relative Absolute Error
RF	Random Forest
UNICEF	United Nations Children's Fund
VIF	Variance Inflation Factor
WHO	World Health Organization

Content

List of Figures

List of Tables

Executive Summary

This report contains Deliverable 3: the results obtained from the application of Machine-Learning (ML) algorithms to identify the main drivers of undernutrition in the Sahel region.

The use of ML models responds to the need to analyse large volumes of data and explore potentially complex and non-linear relationships between multiple factors, representing sociodemographic, environmental, and health conditions. Recent technological and scientific developments have made available new data and more powerful methods that can contribute to deepening our understanding of the conditions that influence undernutrition levels in different countries and regions.

The current study evaluates various AI-machine learning methods using data obtained and pre-processed in previous activities (refer to Deliverables 1 and 2). The aim is to identify the best-performing model for determining the primary drivers of undernutrition among children under five in the Sahel region.

This report details the performed tasks, followed by the outcomes of a test aimed at identifying gaps in the data and assessing collinearity between variables, leading to suitable adjustments to the modelling procedure. Additionally, the report includes details regarding the spatial distribution of undernutrition outcomes in the Sahel region and the correlation between the variables. Also, two ML algorithms - Random Forest (RF) and Boosted Regression Trees (BRT) -, were applied to three undernutrition outcomes separately of children below five years old- stunting, wasting, and underweight. Across the 6 models obtained, the Random Forest algorithm showed a better performance, and it was selected for further analysis, allowing to identification of the most relevant variables influencing the spatial patterns of each undernutrition outcome.

The outputs of the models indicate that women's empowerment, healthcare access, and variations in water supply are common drivers of the different forms of undernutrition, although differences were found among the most relevant drivers for each undernutrition outcome considered. For instance, stunting is associated with technological access, seasonal water availability variation, women's education, contraceptive use, and improved household infrastructure. Wasting outcomes are linked to women's education, decision-making autonomy, antenatal care, and seasonal water variability. Underweight outcomes are influenced by water supply stability, contraceptive use, women's education, decision-making autonomy, and the percentage of grassland.

The results described in Deliverable 3 may contribute to devising a set of recommendations regarding which conditions should be the focus of future undernutrition reduction initiatives and if these should vary by country and sub-region.

1 Introduction

Undernutrition, particularly among children under five is a significant challenge in the Sahel region. Despite considerable commitments in recent decades, and investments in interventions aimed at addressing it, the prevalence remains high (WHO, 2024). Understanding and addressing the complex drivers of undernutrition is crucial for effective programming. The existing literature highlights various factors contributing to undernutrition, including child and maternal characteristics, household conditions, and environmental exposures (Bitew et al., 2021; Fenta et al., 2021b). Nevertheless, the effectiveness of nutrition interventions is often hindered by insufficient consideration of contextual factors and poor program design. A comprehensive contextualized investigation of the drivers of malnutrition has been hard to promote, as research funding bodies frequently assume there is sufficient knowledge to design interventions that address the principal underlying causes of malnutrition, as illustrated by studies on acute malnutrition (Young, 2020)

Moreover, most of the existing studies employed generalized (mixed) linear models for statistical analysis. Despite being widely used for causal inference, these models present limitations such as the allowance of only a small number of covariates and the fact that they do not properly assess multicollinearity (Goldstein et al., 2017; Knol et al., 2008). Furthermore, many studies are limited to cross-sectional designs. To overcome some of these gaps, Machine Learning (ML) methods have recently been suggested to contribute informative data to the study of drivers of undernutrition at the country level and have been adopted, for example, in Ethiopia and Bangladesh (Bitew et al., 2021; Fenta et al., 2021a, 2021b; Mansur et al., 2021). These approaches allow the use of a larger number of predictors/covariates, require fewer assumptions, incorporate 'multi-dimensional correlations', and result in a more flexible relationship among the predictor and outcome variables (Adler et al., 2020; Goldstein et al., 2017; Knol et al., 2008). Nevertheless, so far, such studies are restricted to the intra-country or country-level scope, and mainly address stunting, overlooking other important indicators of child undernutrition such as wasting and micronutrient deficiencies.

This report presents the results of ML algorithms that were applied to the identification of the driving factors of undernutrition in the Sahel region[1]. It describes the implemented procedures of the ML methods, using the available data, which had been previously collected, harmonised, and stored in a geodatabase. To enable the comparison of results across the ML methods tested and identify the best-performing model, the same procedure was applied across the different algorithms. This procedure included the pre-processing of variables, collinearity assessment, model training, validation, performance assessment, and error measurement, as well as the measurement of the relative importance of variables in explaining variations in undernutrition rates.

In the first section, we describe the workflow used for the implementation of the different ML models tested. We note that an initial exploratory analysis of the data was performed, as described in the reports that compose Deliverable 2. We also describe the algorithms used for modelling.

Our preliminary results suggested that the Convolutional Neural Networks (Li et al., 2022) method was not appropriate for use due to the absence of a temporal dimension in the data. These algorithms are often equally or better adapted than 'non-deep' ML algorithms in this case (Capinha et al., 2021). Therefore, two other robust algorithms were explored: Boosted Regression Trees (BRT; Bentéjac et al., 2021) and Random Forest (RF; Parmar et al., 2019).

[1] For the of this study Sahel region boundaries were defined based on spatial coherence and data availability.

In the following section, we report the results of the workflows, including the data pre-processing steps and data gap identification. In addition, collinearity measures were implemented to prevent the inclusion of independent variables with similar variations that were highly correlated. We also describe the ML algorithms used. Each ML algorithm required defining and fine-tuning several parameters with implications for the modelling outputs. A brief description of the parameter setting procedure is provided. This is followed by descriptions of how the models were trained, how their predictive performances were compared, and how the relative importance of each predictor variable was evaluated. Such procedures represent a guarantee of the robustness of the model.

The workflow description is followed by a section detailing the results obtained. It performs a comparative analysis of the results of the models, with particular emphasis on identifying and interpreting the variables identified as most relevant in determining observed variations in levels of undernutrition across the study area.

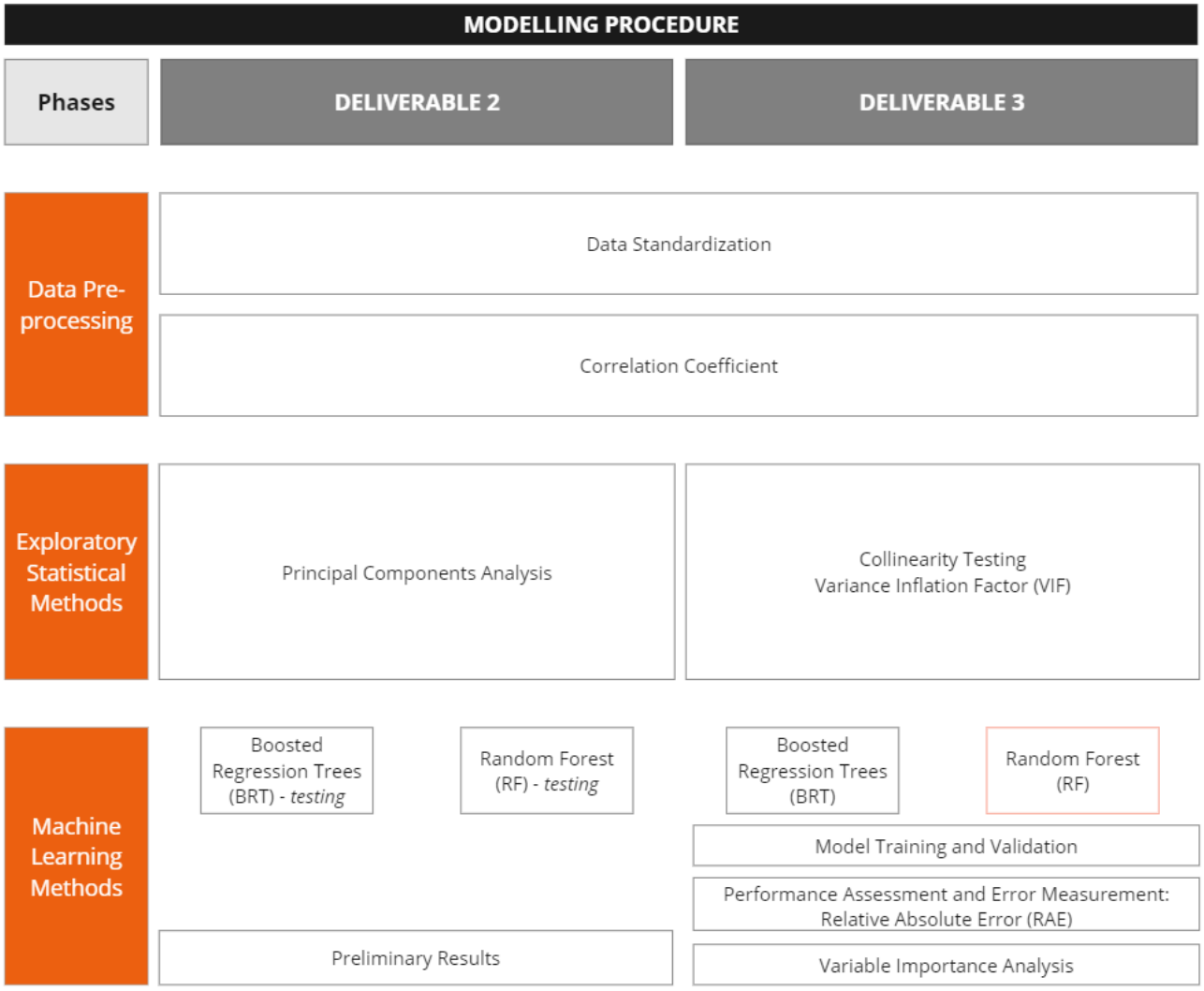

Figure 1.1. Main steps of the modelling procedure

2 Modelling procedure

The modelling procedure was defined according to the objectives, the type of data obtained, the scale of analysis selected, and the requirements of the ML methods applied. Detailed descriptions of the processes of data collection and pre-processing are available in Appendix I.

2.1 Dependent variables

Based on previous literature (UNICEF, 2021; Young, 2020), the available variables considered to best represent the prevalence of undernutrition in the Sahel region were stunting (**Erreur ! Source du renvoi introuvable.**), wasting (Figure 2.**2Erreur ! Source du renvoi introuvable.**), and underweight (Figure 2.**3**), in children under five.

The indicators on undernutrition outcomes were derived from the Demographic and Health Survey (DHS) database and were pre-processed to represent spatial patterns at the subregional level. The methodology employed to obtain these variables is outlined in the reports comprising Deliverable 2.

Each of the variables represents the percentage of children under five suffering from undernutrition in each subregion. Given the diversity of the factors influencing each undernutrition outcome represented by these variables, they were assessed separately, leading to the development of three distinct models: one for stunting (A), another for wasting (B), and a third for underweight (C). The classification thresholds applied in the maps were established based on UNICEF indicators (UNICEF, 2021).

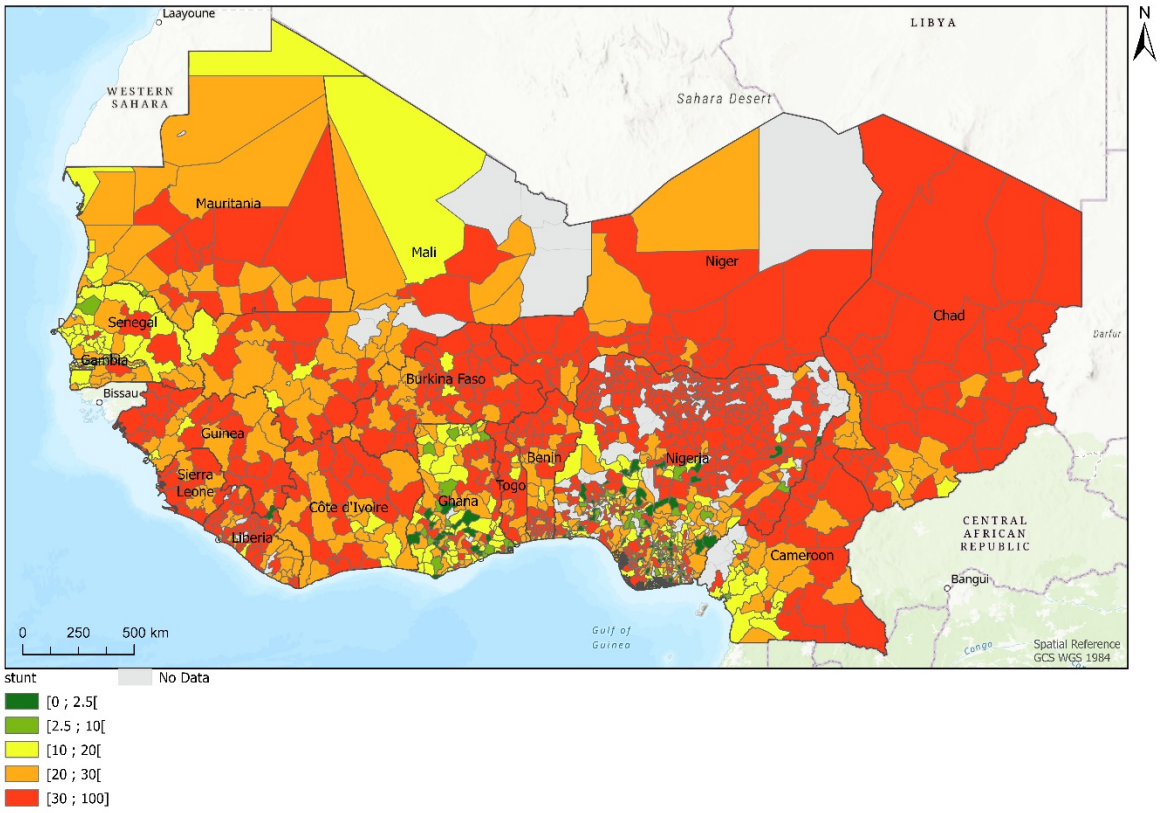

Figure 2.1. Stunting children < 5 years of age (%)

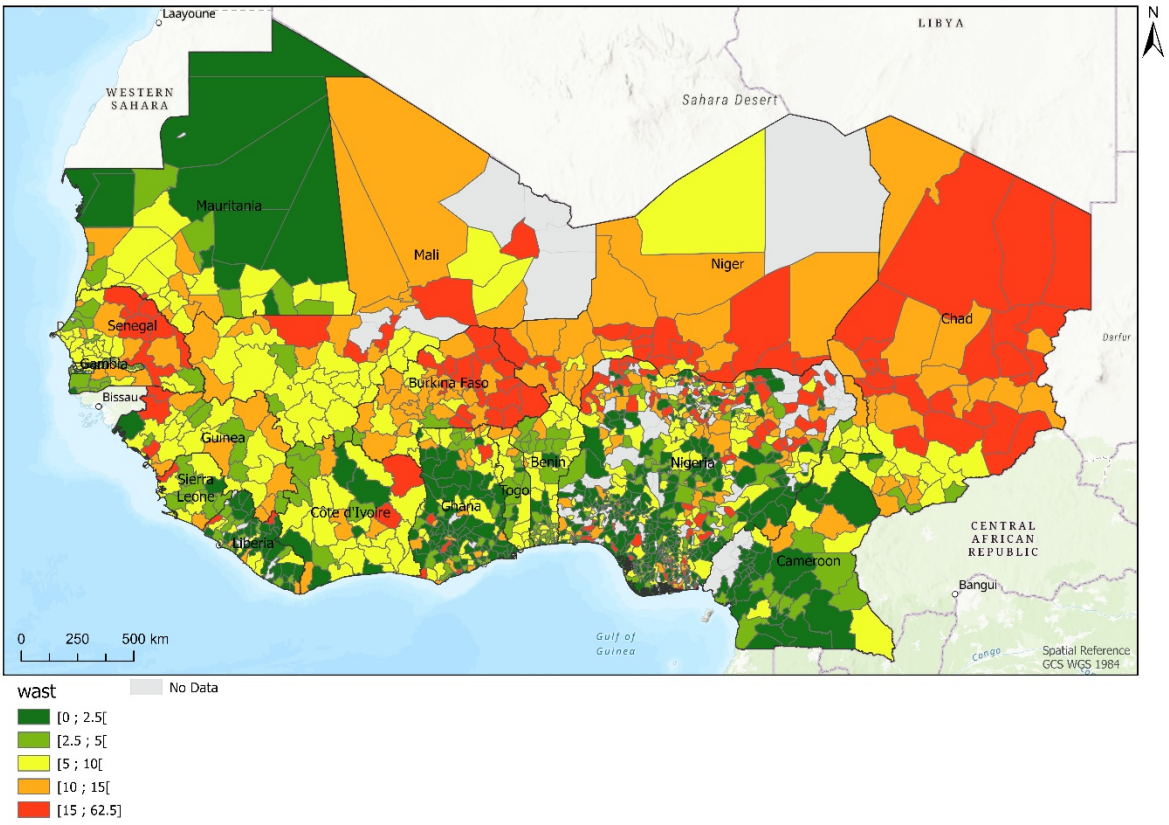

Figure 2.2. Wasting Children <5 years of age (%)

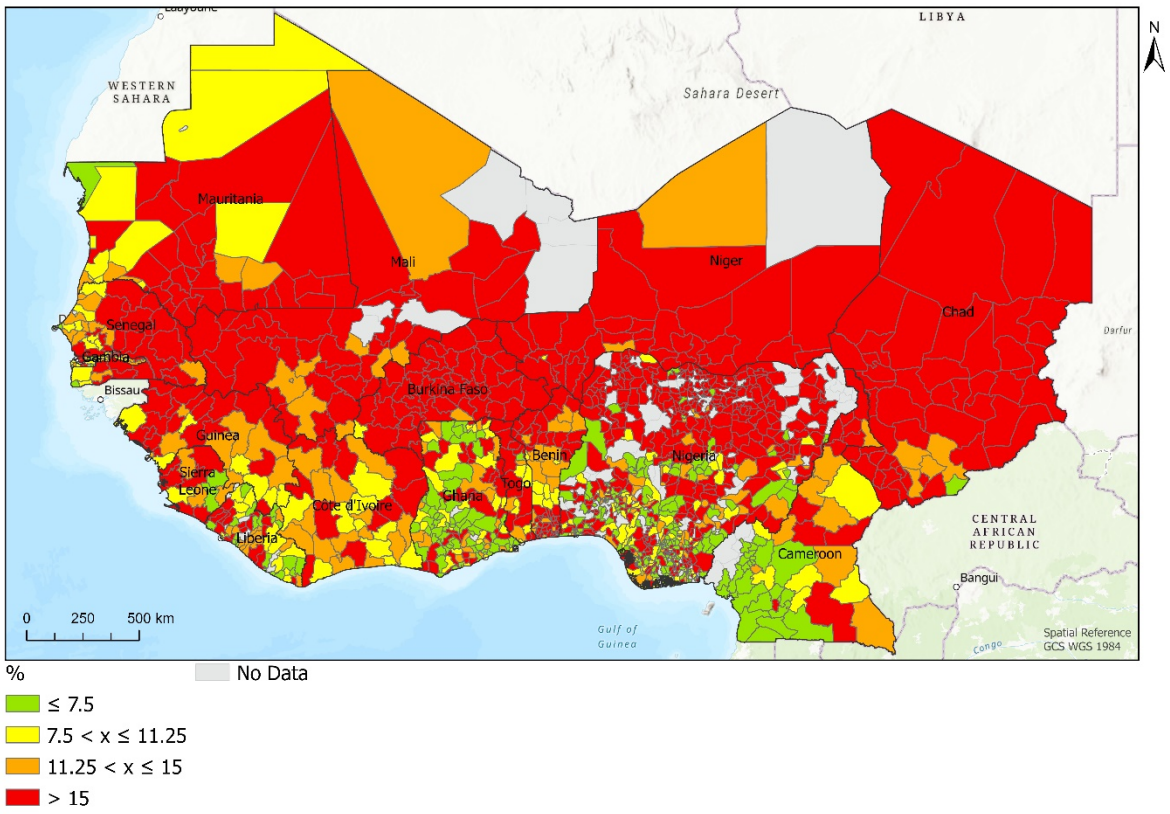

Figure 2.3. Underweight children <5 years of age (%)

The undernutrition outcomes were defined using the recommended threshold of below -2SD (standard deviations) from the median, as defined by the World Health Organization Child Growth Standards.

The mapping of these variables at the sub-regional level showed a relative territorial homogeneity across the countries in the Sahel region, particularly regarding stunting, with most subregions reaching the highest prevalence (>15%). This is not unexpected, since many biophysical and social conditions that influence undernutrition levels are similar in the countries and sub-regions analysed. Following these outcomes, and for mapping and visualization purposes, we calculated the level of severe undernutrition indicators, applying a cut-off, below -3SD from the median, hence focussed this assessment on severe forms of malnutrition.

Using the same classification thresholds as the general undernutrition maps, we present the spatial distribution, at the sub-regional level, of the severe undernutrition outcomes (Figure 2.**4**, 2.5 and 2.6).

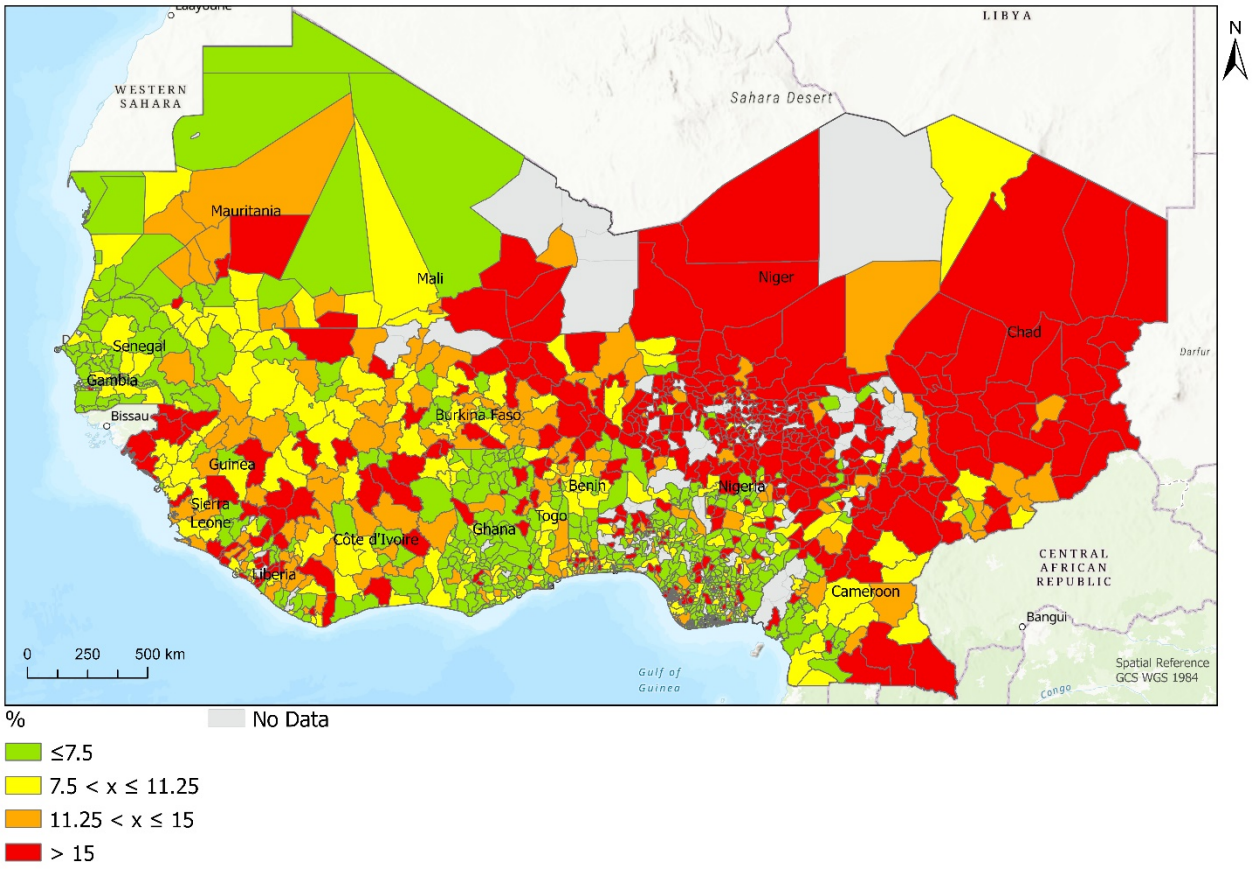

Figure 2.4. Severe stunting children <5 years of age (%)

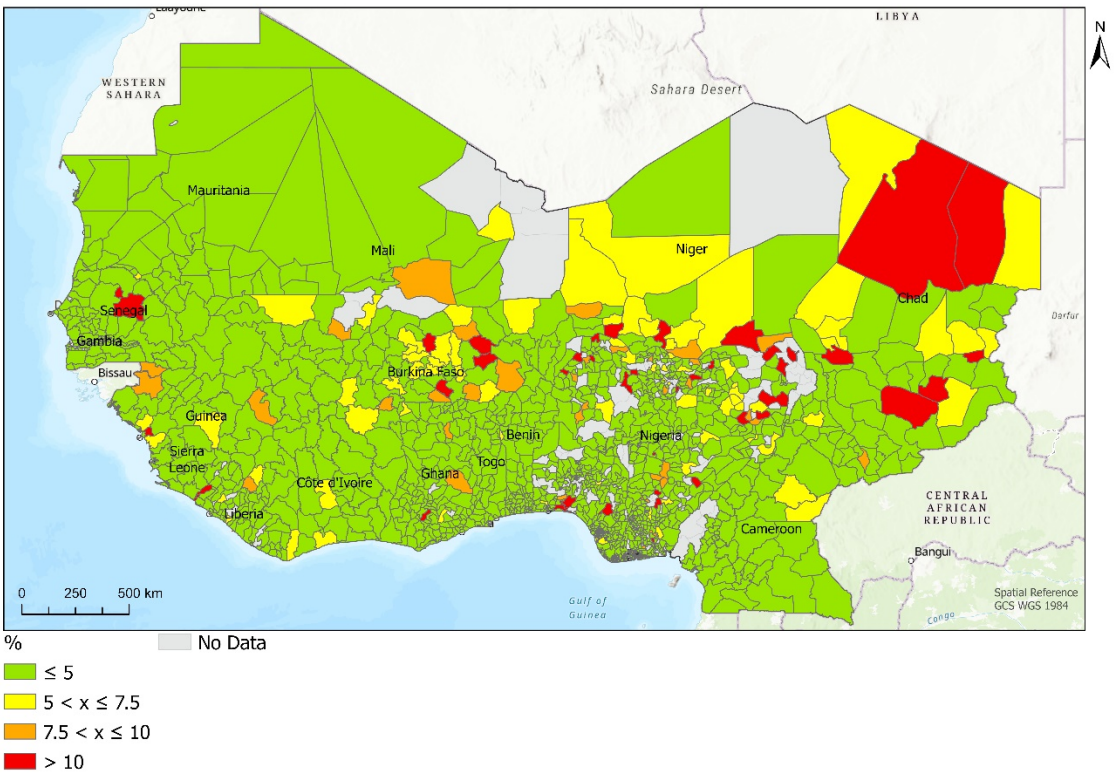

Figure 2.5. Severe wasting children <5 year of age (%)

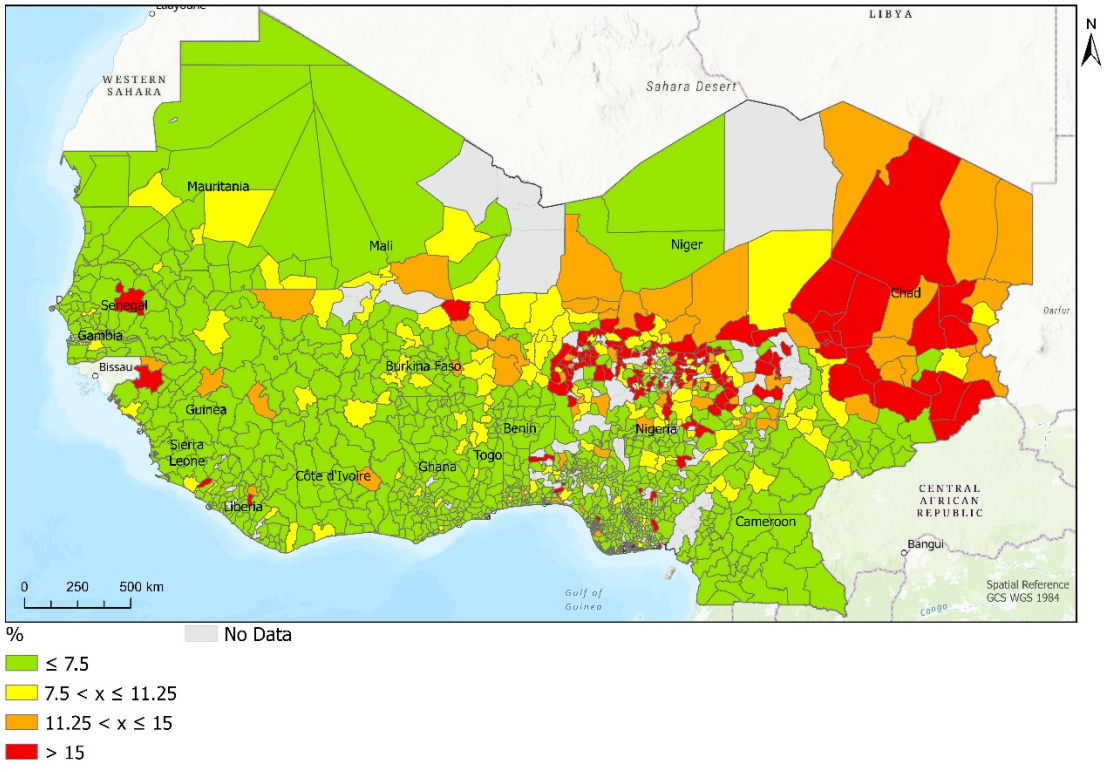

Figure 2.6. Severe underweight children <5 years of age

The distribution of the classes shows greater spatial heterogeneity, for stunting, where the highest class (with values above 15% of stunted children, represented in red) is more dispersed throughout the region, with a cluster in the northeast. In all three maps, most of the subregions classified as having high undernutrition outcomes are found in Chad, Niger, and Nigeria.

2.2 Independent variables

The search for independent variables was guided by the necessity to comprehensively incorporate the impact of various classes of factors: social, demographic, economic, and environmental conditions. This approach was based on the categories of undernutrition drivers identified by UNICEF and supported by relevant literature (Bitew et al., 2022; Chilyabanyama et al., 2022; Fenta et al., 2021a, 2021b; Saroj et al., 2022; UNICEF, 2021; Young, 2020).

The variables obtained were mapped and aggregated at the sub-regional level (administrative level 2). Each variable delineates a particular condition within its respective subregion. For instance, the proportion of women possessing a mobile phone, the percentage of households with access to electricity, the extent of forest land coverage, as well as metrics like mean annual temperature and accumulated precipitation. A total of 90 independent variables were considered, each of which is represented in the table below (Table 2.**1**).

Table 2.1. Comprehensive list of the independent variables explored in the study. Variable name from DHS data indicates the group it represents: ch=children less than five years of age; wm=women; hh = household.

LEVEL	DIMENSION	CATEGORY	INDICATOR	VARIABLE NAME
1. Immediate drivers	Diet Children	Child Dietary Diversity Score	Children < 5 with minimum dietary diversity (%)	ch_mdd
		Child Minimum Acceptable Diet	Children with minimum acceptable diet (%)	ch_mad
		Child Minimum Meal Frequency	Children with minimum meal frequency (%)	ch_mmf
	Health status Children	Child prevalence of anaemia	Children with any anaemiaanaemia (%)	ch_anaemia
		Child disease symptoms	Children with ARI symptoms at any time in the 2 weeks preceding the survey (%)	ch_ari
			Children with fever at any time in the 2 weeks preceding the survey (%)	ch_fever
			Children with diarrhoea at any time in the 2 weeks preceding the survey (%)	ch_diar
	Individual characteristics Children	Sex of child - male	Male children (%)	ch_male
		Sex of child - female	Female children (%)	ch_female
		Child birth weight	Children birth weight < 2.5kg (%)	ch_small
		Child vaccination	Living children aged 12–23 months who received all basic vaccinations (%)	ch_all_vac
			Living children aged 12-23 months with no vaccine received (%)	ch_no_vac
		Childbirth order number	Children 3rd order or more (%)	ch_order3
	Individual characteristics Mothers/women	Birth interval	Women with birth interval < 2 years (%)	wm_birth2
		Mother age at first birth	Women with first birth < 18 years old (%)	wm_birth18
	Health status Mothers/Women	Mother nutritional status (BMI)	Women with BMI < 18.5 kg/m^2 (underweight) (%)	wm_thin
		Prevalence of anaemia in	Women with any anaemia (%)	wm_anaemia

LEVEL	DIMENSION	CATEGORY	INDICATOR	VARIABLE NAME
		women		
2. Underlying drivers	Childcare environment	Age appropriately breastfed - adherence towards WHO optimal feeding recommendations	Children age appropriately breastfed (%)	ch_appbf
	Health services Mothers/Women	Access to antenatal care	Mothers with antenatal care with skilled assistant (%)	wm_skill
		Access to antenatal/prenatal care/ Antenatal visits/Antenatal care 4+ visits	Mothers with antenatal care follow-up with >=4 visits coverage (%)	wm_4visits
		Iron supplementation during pregnancy coverage	Mothers with iron supplementation during pregnancy (%)	wm_iron
		Drugs for intestinal parasites	Mothers who took drugs for intestinal parasites (%)	wm_paras
		Problems accessing health care	Women with at least one problem accessing healthcare, either permission, money, distance or not wanting to go alone (%)	wm_prob
		Knowledge of any contraceptive method	Women with knowledge of at least one contraceptive method (%)	wm_fpknow
		Use of any contraceptive method	Women who currently use any contraceptive method (%)	wm_fpuse
	Health services Children OR mothers/women	Access to health services	Walking only travel time to water (minutes)	tt_wp_wo
			Motorized travel time to water (minutes)	tt_wp_mo
3. Systemic drivers	Individual characteristics Women	Marital status	Women who are married or in a union (%)	wm_union
		Literacy	Women literacy - higher than secondary or can read part or whole sentence (%)	wm_liter
		Occupation	Women currently employed (%)	wm_employ
			Women with agriculture/manual occupation (%)	wm_agriman
		Media exposure	Women who access all the three media (TV, radio, newspaper) on a weekly basis (%)	wm_media3
		Media exposure	Women who access none of the three media (tv, radio, newspaper) at least once a week (%)	wm_nomedia
	Women empowerment	Ownership of assets	Women who use a bank account (%)	wm_bank
		Ownership of assets	Women who own a house, either alone or jointly (%)	wm_house
		Ownership of assets	Women who own a land, either alone or jointly (%)	wm_land
		Ownership of assets	Women who own a mobile phone, either alone or jointly (%)	wm_mobile
		Decision-making	Women who decide on health, purchases and visits, either alone or jointly with partner (%)	wm_decide
	Religion	Religion	Women affiliated with Islam (%)	wm_muslim
			Women affiliated with Christianity (%)	wm_christian
			Women affiliated with traditional religions	wm_tradit

LEVEL	DIMENSION	CATEGORY	INDICATOR	VARIABLE NAME
			(%)	
			Women affiliated with other/no religion (%)	wm_no_relig
	Living environment	Access to electricity	Households with electricity (%)	hh_electric
		Handwashing facility	Households with basic handwashing facility (%)	hh_basic
			Households with limited handwashing facility (%)	hh_limited
		Type of cooking fuel	Households using solid fuel for cooking (%)	hh_solid_fuel
			Households using clean fuel for cooking (%)	hh_clean_fuel
	Livelihood systems - socioeconomic status	Socioeconomic status	People in the first or second wealth quintile - poor (%)	hh_poor
	Livelihood systems - Household assets	Household assets	Households with TV (%)	hh_tv
			Households with radio (%)	hh_radio
			Households with mobile or telephone (%)	hh_phone
			Households with refrigerator (%)	hh_fridge
			Households with car or motorcycle (%)	hh_motocar
			Households with agricultural land (%)	hh_agriland
			Households with livestock (%)	hh_livestock
	Livelihood systems & resources	Access to health services	Walking only travel time to healthcare facility (minutes)	tt_hc_wo
			Motorized travel time to healthcare facility (min)	tt_hc_mo
	Governance & Political Economy	Poverty	Deprivation index	grdi
			Unimproved or no drinking water (% population)	udw
			Unimproved or no sanitation (% population)	usa
		Socio demography	Population density (n° people/km²)	pop_dens
		Civil Insecurity	Nr. victims from conflicts per year over 21 years (n°/year)	fat_p_year
			Nr. Conflicts per year over the last 21 years (n°/year)	event_freq
	Food systems	Crop production	Production of agricultural crops (ton)	prod_total
			Production/km² (ton/km² farming area)	agr_prod
		Access to irrigation (household)/ Low water availability	Seasonal water supply variability (index)	sev
		Crop growing season	Growing season length (nr. days)	gsl
			Mean temperature of the growing season (ºC)	gst
	Environment & seasonality	Climate	Mean annual air temperature (ºC)	mat
			Annual precipitation amount (kg/m²)	ap
			Maximum monthly potential evapotranspiration (kg/m²)	max_po_ev
			Minimum monthly potential evapotranspiration (kg/m²)	min_po_ev

LEVEL	DIMENSION	CATEGORY	INDICATOR	VARIABLE NAME
			Mean monthly potential evapotranspiration (kg/m²)	mean_po_ev
		Topography	Mean altitude (m)	DEM
		Land Use & Landcover	Net primary productivity (g carbon /m²)	npp
			Agriculture (% area)	agr_per
			Grassland, lichen mosses or sparse vegetation (%)	grass_per
			Forest + shrub (wildland) (% area)	forest_per
			Urbanized (% area)	urban_per
			Water bodies (% area)	water_per
			Wetlands (% area)	wetlan_per
			Rocks and sparsely vegetated (% area)	bare_per
		Human Intervention	Road density	road_dens
			Human Modification Index	hmts
		Extreme events	Arid (% of area)	arid_per
			low drought risk (% area)	low_per
			low/medium drought risk (% area)	lo_me_per
			medium drought risk (% area)	med_per
			medium/high drought risk (% area)	me_hi_per
			high drought risk (% area)	high_per

The analyses were conducted for 16 countries within the Sahel region, comprising a total of 1362 subregions, originally chosen from a larger pool of 1528 (Figure 2.**7**). 166 subregions were excluded from the analysis, due to data unavailability from the Demographic and Health Survey (DHS). This unavailability was due to two primary reasons: 1) security concerns, leading to the absence of DHS surveys in certain sub-regions, as detailed in each survey and country report, and 2) data compilation errors, wherein available survey data clusters lacked correct georeferencing information (i.e., no attributed coordinates), making it impossible to establish spatial connections with the subregions. Further detail is provided in Appendix II.

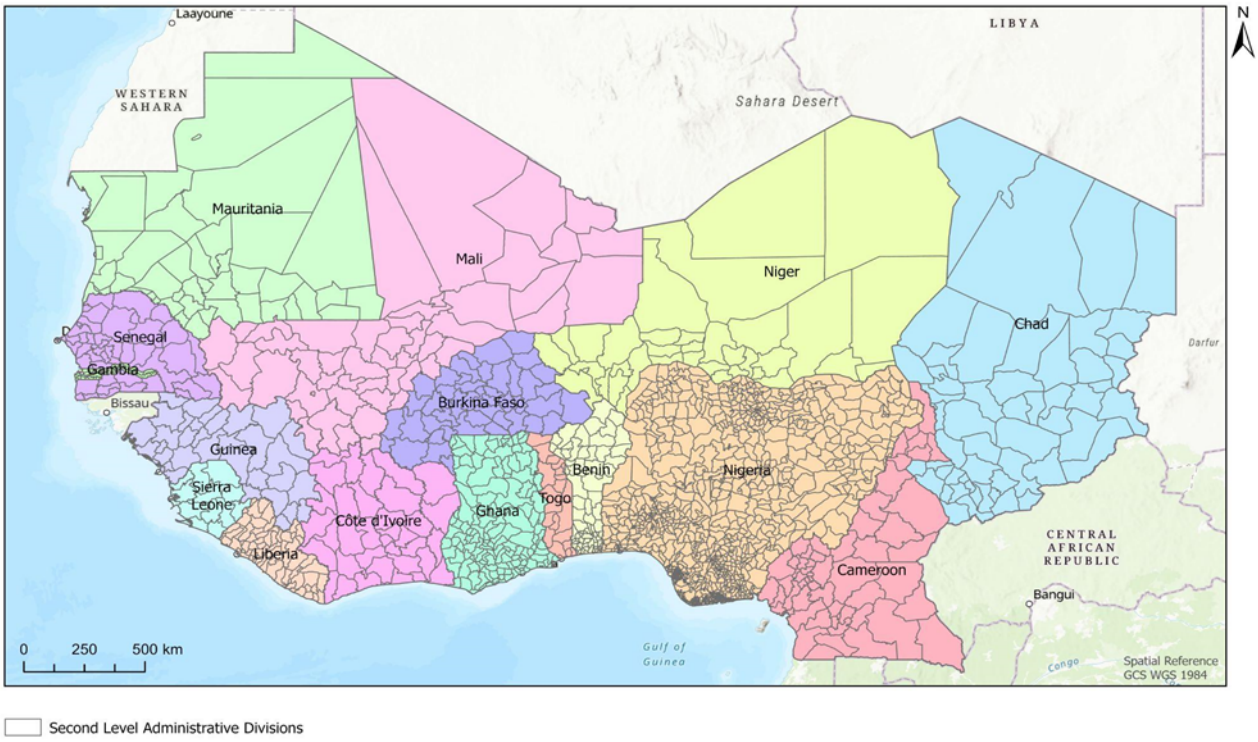

Figure 2.7. Map of the study area and its subregions.

2.3 Model Development and Evaluation

2.3.1 Collinearity testing

Prior to model implementation, the independent variables underwent a collinearity testing using the Variance Inflation Factor (VIF). This metric quantifies the extent to which the variance of an independent variable is inflated by its correlation with multiple other predictor variables. To prevent redundancy in the information conveyed by predictors, or overfitting of the models, the variables that had a VIF threshold >10 were excluded from further processing, as they were considered problematic due to multicollinearity (Curto & Pinto, 2011).

Furthermore, to understand the relation between specific variables, we calculated the Pearson correlation coefficient (PCC) for each pair of independent variables. More details can be found in Deliverable 2. A graph with a visual representation of the association between each pair of variables is presented in Appendix 3. The red circles show a negative association, and blue circles a positive association, meaning that the values of both variables follow the same direction.

2.3.2 Machine-learning methods

The ML approach uses algorithms to learn the relationships between the response and predictors variables, based on consistent patterns in the data. These models can identify simple to high-dimensional, linear and nonlinear, statistical relations between variables.

The ML methods applied here are based on ensembles of regression trees, namely Boosted Regression Trees (BRT) and Random Forest (RF), which have been previously applied for undernutrition research (Chilyabanyama et al., 2022; Fenta et al., 2021a, 2021b).

A) Boosted Regression Trees (BRT)

The BRT is a self-learning decision tree-based algorithm that combines a large number of tree models in sequence to optimize predictive performance. The boosting approach fits the model iteratively to the training data. The first regression tree reduces the loss function, while the following is fitted to the residuals. The procedure is repeated until a prespecified criterion is met (e.g., a specific number of trees, or minimal residual values). During the boosting process, the construction of each tree depends on the results obtained by the chain of previous trees, with a focus on rectifying misclassified samples (Elith et al., 2008).

Before calibrating BRT models, it is typical to define four parameters: 1) learning rate (LR), 2) tree complexity (TC), 3) number of trees (NT) and (4) bag fraction (BF). In accordance with previous recommendations (Elith et al., 2008), we set the LR to a low value of 0.001, TC to an 'intermediate' value of 5, and BF to 66. The number of trees was determined automatically using the ‘gbm.step’ function from the ‘dismo’ package in R (Hijmans et al., 2022).

B) Random Forest (RF)

RF is a non-parametric ensemble ML model that builds multiple uncorrelated decision trees (classifiers). Each tree is constructed using a randomly chosen subset of the original samples, estimating the mean prediction of the individual trees. The algorithm combines all these predictions and identifies the optimal set of variables. RF shows very good accuracy, high flexibility for computing several types of statistical data analysis, an ability to deal with missing values, outliers, and mislabelled samples (Cutler et al., 2007). It performs both classification and regression, and mitigates overfitting (Breiman, 2001; Wei et al., 2019). The challenges are related to parameter selection, interpretation of results, and computation time (Wei et al., 2019).

RF requires the user to define the number of trees to create and the number of variables to use in each split. The number of trees was defined as 5,000 and the number of variables was set at either 11, 22, or 44, differing from fold to fold and from model to model, based on the function ‘tuneRF’ from the ‘RandomForest package’ (Breiman, 2001). When creating a tree, RF sets aside one-third of the data – called ‘out-of-bag’ (OOB) - which is used for prediction and calculating error rates as an additional model validation tool. The variable importance metric used was the mean decrease in accuracy, which measures how much the model loses accuracy when removing each variable.

2.3.3 Model training and validation

Following the removal of the collinear variables, a data partitioning technique was applied to establish separate samples for model training and validation. This approach allowed for model performance assessment using data that wasn’t involved in the training phase. For this purpose, a k-fold cross-validation technique was implemented, randomly dividing the dataset into 5 groups (5-folds), with each containing 20% of the data (i.e., 20% of the subregions). The models were run five times, with each execution using 4 folds for training and 1 fold left for validation. All subregions were included in both training and testing stages. The final results represent the mean of all folds.

2.3.4 Performance assessment and error measurement

To measure and compare the predictive performance of the models, we used the relative absolute error (RAE).

RAE is expressed as a ratio and compares the mean of deviations between predicted and observed values, and the mean of errors produced by a naive model that consists simply of the mean of observed values in the validation fold. The lower the RAE value, the higher the performance of the model. A value of 1 corresponds to the error level obtained for the naive model and a value of 0 corresponds to a perfect match between predicted and observed values. To enable interpreting the importance of the variables as main drivers, the model showing the lowest RAE was selected.

2.3.5 Variable importance

The importance of the variables in the models was assessed based on the percentage of the variance of the dependent variable (undernutrition outcome) explained by each independent variable.

A threshold of 75% explained variance was defined to select the most influential drivers for each undernutrition outcome, leading to the selection of the variables that met this criterion.

3 Results

3.1 Collinearity - Variance Inflation Factor and pairwise Pearson correlation

The results obtained from the VIF were critically examined and multiple runs were conducted to select the variables. Despite the threshold of VIF >10, two variables with VIF values slightly higher were retained, due to their potential relevance for understanding undernutrition: 'People in the First or Second Lowest Wealth Quintile - Poor (%)' (hh_poor) and 'Women Affiliated with Traditional Religions (%)' (wm_trad).

Ultimately, a total of 20 variables with VIF values exceeding the threshold of 10 were excluded due to their high collinearity (Table 3.1).

Table 3.1. Excluded variables due to high collinearity, determined by VIF results (VIF>10)

INDICATOR	VARIABLE NAME
Households with TV (%)	hh_tv
Women affiliated with Islam (%)	wm_muslim
Women affiliated with Christianity (%)	wm_christian
Unimproved or no sanitation (% population)	usa
Walking only - travel time to water (mins.)	tt_wp_wo
Motorized travel time to healthcare (mins.)	tt_hc_mo
Walking only travel time to healthcare (mins.)	tt_hc_wo
Deprivation index	grdi
Agriculture (% area)	agr_per
Forest or shrub (wildland) (% area)	forest_per
Urbanized (% area)	urban_per
Net primary productivity (g carbon /m^2)	npp
Mean altitude (m)	DEM
Minimum monthly potential evapotranspiration (kg/m^2)	min_po_ev
Mean monthly potential evapotranspiration (kg/m^2)	mean_po_ev
Maximum monthly potential evapotranspiration (kg/m^2)	max_po_ev
Mean annual air temperature (°C)	mat
Annual precipitation (kg/m^2)	ap
Growing season length (n° of days)	gsl
Mean temperature of the growing season (°C)	gst

As a result of the previous steps, 66 out of the initial 90 independent variables were included in the modelling procedure. The list of these variables is presented below (Table 3.2).

Table 3.2. Selected variables for modelling

LEVEL	DIMENSION	CATEGORY	INDICATOR	VARIABLE NAME
1. Immediate drivers	Diet Children	Child Dietary Diversity Score	Children < 5 with minimum dietary diversity (%)	ch_mdd
		Child Minimum Acceptable Diet	Children with minimum acceptable diet (%)	ch_mad
		Child Minimum Meal Frequency	Children with minimum meal frequency (%)	ch_mmf
	Health status Children	Child prevalence of anaemia	Children with any anaemia (%)	ch_anaemia
		Child disease symptoms	Children with ARI symptoms at any time in the 2 weeks preceding the survey (%)	ch_ari
			Children with fever at any time in the 2 weeks preceding the survey (%)	ch_fever
			Children with diarrhoea at any time in the 2 weeks preceding the survey (%)	ch_diar
	Individual characteristics Child	Sex of child - female	Female children (%)	ch_female
		Child birth weight	Children birth weight < 2.5kg (%)	ch_small
		Child Vaccination	Living children aged 12–23 months who received all basic vaccinations (%)	ch_all_vac
			Living children aged 12-23 months with no vaccine received (%)	ch_no_vac
		Childbirth order number	Children 3rd order or more (%)	ch_order3
	Individual characteristics Mothers/women	Birth interval	Women with birth interval < 2 years (%)	wm_birth2
		Mother age at first birth	Women with first birth < 18 years old (%)	wm_birth18
	Health status Mothers/women	Mother nutritional status (BMI)	Women with BMI < 18.5 kg/m2 (underweight) (%)	wm_thin
		Prevalence of anaemia in women	Women with any anaemia (%)	wm_anaemia
2. Underlying drivers	Childcare environment	Age appropriately breastfed - adherence towards WHO optimal feeding recommendations	Children age appropriately breastfed (%)	ch_appbf
	Health services Mothers/women	Access to antenatal care	Mothers with antenatal care with skilled assistant (%)	wm_skill
		Access to antenatal/prenatal care/ Antenatal visits/Antenatal care 4+ visits	Mothers with antenatal care follow-up with >=4 visits coverage (%)	wm_4visits
		Iron supplementation during pregnancy coverage	Mothers with iron supplementation during pregnancy (%)	wm_iron
		Drugs for intestinal parasites	Mothers who took drugs for intestinal parasites (%)	wm_paras
		Problems accessing health care	Women with at least one problem accessing healthcare, either permission, money, distance or not wanting to go alone (%)	wm_prob
		Knowledge of any contraceptive method	Women with knowledge of at least one contraceptive method (%)	wm_fpknow
		Use of any contraceptive method	Women who currently use any contraceptive method (%)	wm_fpuse
3. Systemic	Individual	Marital status	Women who are married or with a partner a union	wm_union

LEVEL	DIMENSION	CATEGORY	INDICATOR	VARIABLE NAME
drivers	characteristics Women		(%)	
		Literacy	Women literacy - higher than secondary or can read part or whole sentence (%)	wm_liter
		Occupation	Women currently employed (%)	wm_employ
			Women with agriculture/manual occupation (%)	wm_agriman
		Media exposure	Women who access all the three media (TV, radio, newspaper) on a weekly basis (%)	wm_media3
		Media exposure	Women who access none of the three media (tv, radio, newspaper) at least once a week (%)	wm_nomedia
	Women empowerment	Ownership of assets	Women who use a bank account (%)	wm_bank
		Ownership of assets	Women who own a house, either alone or jointly (%)	wm_house
		Ownership of assets	Women who own a land, either alone or jointly (%)	wm_land
		Ownership of assets	Women who own a mobile phone, either alone or jointly (%)	wm_mobile
		Decision-making	Women who decide on health, purchases and visits, either alone or jointly with partner (%)	wm_decide
	Religion	Religion	Women affiliated with traditional religions (%)	wm_tradit
	Living environment	Access to electricity	Households with electricity (%)	hh_electric
		Handwashing facility	Households with basic hand washing facility (%)	hh_basic
			Households with limited hand washing facility (%)	hh_limited
		Type of cooking fuel	Households using solid fuel for cooking (%)	hh_solid_fuel
			Households using clean fuel for cooking (%)	hh_clean_fuel
	Livelihood systems - socioeconomic status	Socioeconomic status	People in the first or second wealth quintile - poor (%)	hh_poor
	Livelihood systems - Household assets	Household assets	Households with radio (%)	hh_radio
			Households with mobile or telephone (%)	hh_phone
			Households with refrigerator (%)	hh_fridge
			Households with car or motorcycle (%)	hh_motocar
			Households with agricultural land (%)	hh_agriland
			Households with livestock (%)	hh_livestock
	Livelihood systems & resources	Access to water	Motorized travel time to water (minutes)	tt_wp_mo
		Poverty	Unimproved or no drinking water (% population)	udw
		Sociodemographics	Population density (n° people/km²)	pop_dens
		Civil Insecurity	Nr. victims from conflicts per year over 21 years (n°/year)	fat_p_year
			Nr. Conflicts per year over the last 21 years (n°/year)	event_freq
	Food systems	Crop production	Production of agricultural crops (ton)	prod_total
			Production/km² (ton/km² farming area)	agr_prod
		Access to irrigation (household)/ low water availability	Seasonal water supply variability (index)	sev
		Land use & landcover	Grassland, lichen mosses or sparse vegetation (%)	grass_per
			Water bodies (% area)	water_per
			Rocks and sparsely vegetated (% area)	bare_per
	Environment & seasonality	Human intervention	Road density	road_dens
		Human intervention	Human Modification Index	hmts
		Extreme events	Arid (% of area)	arid_per
		Extreme events	Low drought risk (% area)	low_per
		Extreme events	Low/medium drought risk (% area)	lo_me_per
		Extreme events	Medium drought risk (% area)	med_per
		Extreme events	Medium/high drought risk (% area)	me_hi_per

For further understanding of the relationship between variables, we have also computed a pairwise correlation, using PCC.

The results show that women being thin, having their first birth before 18 years old, being married or in a union, lacking of access to media, and the household relying on solid fuel, are variables associated with an increased % of undernourished children. On the contrary, the higher the % of literate women, the lower the % of undernourished children.

Women's literacy exhibits negative correlation with women being thin, having their first child earlier than 18 years old. On the other hand, it shows a positive correlation with the % of households with possessions (fridge, tv, radio, electricity, clean fuel), with mobile phone or bank account ownership, with access to media and the possibility to make decisions, and access to antenatal care and visits.

Households with basic hand washing facilities (hh_basic) decrease when the proportion of poor people (hh_poor) and households using solid fuel (hh_solid) increase. This variable is also negatively correlated with households owning livestock (hh_livestock), the deprivation index (grdi) and with women whose first birth was before 18 years old (wm_birth18). When the % of households using solid fuel increases, the proportion of households having electricity, tv, phone, fridge, or radio decreases.

3.2 Model assessment - Performance and error measurement

Both methods - BRT and RF - were applied to each undernutrition outcome (stunting, wasting and underweight) and showed a reasonable performance level. The relative absolute error indicated that the RF model showed an equal or slightly better performance than BRT for all outcomes. The undernutrition indicator predicted with the highest accuracy was stunting (min. RAE = 0.65), followed by underweight (min RAE = 0.71) and finally wasting showed the highest error rate (min RAE = 0.92) (Table 3.3).

Table 3.3. Performance and error measurement of the two models tested (BRT-Boosted Regression Trees and RF-Random Forest) for Stunting, Wasting, and Underweight

MODEL	BRT - RELATIVE ABSOLUTE ERROR	RF - RELATIVE ABSOLUTE ERROR
Stunting	0.68	0.65
Wasting	0.92	0.92
Underweight	0.73	0.71

3.3 Variable importance - drivers of undernutrition

According to the model performance assessment applied, the RF method showed a lower RAE and higher accuracy. As such, the identification of the most important variables was based on the variable importance metrics obtained for this algorithm only.

Tables 3.4, 3.5, 3.6 and 3.7 show the most influential factors in the RF models for each of the three dependent variables. The variable factors represented explain a cumulative variance of the dependent variable of at least 75%. For the stunting model, 18 variables were selected, for the underweight model 20 variables, and the wasting model required 27 variables to explain 75% of cumulative variance.

Table 3.4. Variable ranking for stunting, wasting, and underweight based on the results obtained with the Random Forest model

VARIABLES RANKING	STUNTING	WASTING	UNDERWEIGHT
1	Women who own a mobile phone, either alone or jointly (%)	Women literacy - higher than secondary or can read part or whole sentence (%)	Seasonal water supply variability (index)
2	Seasonal water supply variability (index)	Women who decide on health, purchases, and visits, either alone or jointly with partner (%)	Women who currently use any contraceptive method (%)
3	Women literacy - higher than secondary or can read part or whole sentence (%)	Mothers with antenatal care follow-up with ≥ 4 visits coverage (%)	Women literacy - higher than secondary or can read part or whole sentence (%)
4	Women who currently use any contraceptive method (%)	Mothers with antenatal care with skilled assistant (%)	Women who decide on health, purchases and visits, either alone or jointly with partner (%)
5	Households with refrigerator (%)	Seasonal water supply variability (index)	Grassland, lichen mosses or sparse vegetation (%)
6	People in the first or second wealth quintile - poor (%)	Grassland, lichen mosses or sparse vegetation (%)	Mothers with antenatal care with skilled assistant (%)
7	Women with first birth < 18 years old (%)	Children 3rd order or more (%)	Women who own a mobile phone, either alone or jointly (%)
8	Women who use a bank account (%)	Production/km^2 (ton/km^2 farming area)	Women with first birth < 18 years old (%)
9	Women who decide on health, purchases and visits, either alone or jointly with partner (%)	Households with livestock (%)	Women who are married or in a union (%)
10	Women who own a house, either alone or jointly (%)	Living children aged 12-23 months with no vaccine received (%)	People in the first or second wealth quintile - poor (%)
11	Human Modification Index	Households using solid fuel for cooking (%)	Women with BMI < 18.5 mg/m^2(underweight) (%)
12	Women who access none of the three media (tv, radio, newspaper) at least once a week (%)	Women who are married or in a union (%)	Mothers with antenatal care follow-up with ≥4 visits coverage (%)
13	Grassland, lichen mosses or sparse vegetation (%)	Women who use a bank account (%)	Mothers with iron supplementation during pregnancy (%)
14	Population density (nr. people/km^2)	Women currently employed (%)	Women who own a house, either alone or jointly (%)
15	Mothers with antenatal care with skilled assistant (%)	Women who currently use any contraceptive method (%)	Living children aged 12-23 months with no vaccine received (%)

VARIABLES RANKING	STUNTING	WASTING	UNDERWEIGHT
16	Women who are married or in a union (%)	Arid land (% of area)	Women who access none of the three media (TV, radio, newspaper) at least once a week (%)
17	Living children aged 12-23 months with no vaccine received (%)	Road density	Women who use a bank account (%)
18	Living children aged 12–23 months who received all basic vaccinations (%)	Households with limited hand washing facility (%)	Women with birth interval < 2 years (%)
19		Women who access all the three media (TV, radio, newspaper) on a weekly basis (%)	Households with refrigerator (%)
20		Households with radio (%)	Living children aged 12–23 months who received all basic vaccinations (%)
21		Women who access none of the three media (TV, radio, newspaper) at least once a week (%)	
22		People in the first or second wealth quintile - poor (%)	
23		Women with first birth < 18 years old (%)	
24		Households with agricultural land (%)	
25		Households with refrigerator (%)	
26		Children with any anaemia (%)	
27		Households using clean fuel for cooking (%)	

Table 3.5. Variable ranking for stunting, variance (%), cumulative variance (%) and expected effect on undernutrition outcomes, obtained from correlation levels between variables

VARIABLES RANKING	VARIABLE NAME	VARIANCE (%)	CUMULATIVE VARIANCE (%)	ASSOCIATION WITH UNDERNUTRITION LEVELS
1	wm_mobile	11.61	11.61	↘
2	sev	9.90	21.51	↗
3	wm_liter	5.84	27.35	↘
4	wm_fpuse	5.75	33.10	↘
5	hh_fridge	5.30	38.41	↘

6	hh_poor	4.28	42.69	↗
7	wm_birth18	4.16	46.84	↗
8	wm_bank	3.79	50.63	↘
9	wm_decide	3.44	54.07	↘
10	wm_house	3.16	57.23	↘
11	hmts	2.96	60.19	↘
12	wm_nomedia	2.80	62.99	↗
13	grass_per	2.69	65.68	↗
14	pop_dens	2.29	67.97	↘
15	wm_skill	1.90	69.87	↘
16	wm_union	1.88	71.75	↘
17	ch_no_vac	1.87	73.62	↗
18	ch_all_vac	1.72	75.34	↘

↗ Expected to increase undernutrition levels

↘ Expected to decrease undernutrition levels

Table 3.6. Variable Ranking for wasting, variance (%), cumulative variance (%) and expected effect on undernutrition outcomes, obtained from correlation levels between variables

VARIABLES RANKING	VARIABLE NAME	VARIANCE (%)	CUMULATIVE VARIANCE (%)	EXPECTED EFFECT ON UNDERNUTRITION LEVELS
1	wm_liter	6.16	6.16	↘
2	wm_decide	4.63	10.78	↘
3	wm_4visits	4.41	15.20	↘
4	wm_skill	4.28	19.48	↘
5	sev	3.78	23.26	↗
6	grass_per	3.49	26.75	↘
7	ch_order3	3.40	30.14	↗
8	agr_prod	3.18	33.32	↘
9	hh_livestock	3.14	36.46	↘
10	ch_no_vac	3.09	39.55	↗
11	hh_solid_fuel	2.82	42.36	↗
12	wm_union	2.80	45.17	↘
13	wm_bank	2.78	47.95	↘
14	wm_employ	2.75	50.70	↘
15	wm_fpuse	2.71	53.40	↘
16	arid_per	2.63	56.03	↗
17	road_dens	2.49	58.52	↘
18	hh_limited	2.30	60.83	↗
19	wm_media3	2.14	62.96	↘
20	hh_radio	2.06	65.02	↘
21	wm_nomedia	2.01	67.03	↗
22	hh_poor	1.60	68.63	↗
23	wm_birth18	1.46	70.08	↗
24	hh_agriland	1.44	71.53	↘
25	hh_fridge	1.42	72.95	↘
26	ch_anaemia	1.35	74.30	↗
27	hh_clean_fuel	1.27	75.57	↘

↗ Expected to increase undernutrition levels
↘ Expected to decrease undernutrition levels

Table 3.7. Variable Ranking for underweight, variance (%), cumulative variance (%) and expected effect on undernutrition outcomes, obtained from correlation levels between variables

VARIABLES RANKING	VARIABLE NAME	VARIANCE (%)	CUMULATIVE VARIANCE (%)	EXPECTED EFFECT ON UNDERNUTRITION LEVELS
1	sev	9.13	9.13	↗
2	wm_fpuse	8.78	17.92	↘
3	wm_liter	6.22	24.14	↘
4	wm_decide	5.74	29.88	↘
5	grass_per	5.57	35.46	↘
6	wm_skill	4.97	40.43	↘
7	wm_mobile	4.75	45.18	↘
8	wm_birth18	3.47	48.65	↗
9	wm_union	3.06	51.71	↘
10	hh_poor	2.84	54.55	↗
11	wm_thin	2.79	57.34	↗
12	wm_4visits	2.52	59.86	↘
13	wm_iron	2.12	61.98	↘
14	wm_house	2.02	64.00	↘
15	ch_no_vac	1.99	65.99	↗
16	wm_nomedia	1.89	67.88	↗
17	wm_bank	1.82	69.69	↘
18	wm_birth2	1.69	71.39	↗
19	hh_fridge	1.65	73.03	↘
20	ch_all_vac	1.56	74.60	↘

↗ Expected to increase undernutrition levels
↘ Expected to decrease undernutrition levels

3.3.1 Stunting

For the stunting model, the top 5 most important variables highlight the relevance of technological access (wm_mobile), seasonal water availability variation (sev), women's education (wm_liter), contraceptive use (wm_fpuse), and improved household infrastructure (hh_fridge) in addressing stunting in children.

The most important variable for this model is 'women who own a mobile phone, either alone or jointly (%)', which may translate the various benefits of mobile phone ownership, including access to information (e.g. health-related information), access to healthcare services and participation in digital health programs, which provide opportunities for women's engagement in the health system. Also, mobile phones promote social support, economic opportunities, and behaviour change interventions. Moreover, in many low and middle-income countries, where gender inequality persists, the ownership of mobile phones can act as an equalizer, providing women with the same information as their male counterparts (LeFevre et al., 2020), acting as an important tool for women's empowerment.

The second most important variable in this model - 'sev - seasonal water supply variability' - suggests that seasonal variations in water supply may have an important impact on stunting outcomes. The importance of variations in water supply can be attributed to several reasons, such as insufficient access to clean and safe drinking water, which may compromise adequate water intake, sanitation, and hygiene, raising the

risk of disease transmission and malnutrition (e.g. repeated diarrhoea, infections, and environmental enteric dysfunction). Water scarcity can also have an impact on food production, contributing to nutritional deficiencies. Furthermore, discrepancies in water supply are commonly associated with socioeconomic inequities. All these aspects indirectly affect growth, immunity and physical and neurodevelopment of children under five (Ogunniyi et al., 2021; Woldesenbet at al., 2023).

The third most important, 'women's education (wm_liter)' variable indicates that higher literacy rates among women were found to be associated with lower rates of stunting. This may be due to the relationship between education and empowerment in women, reflecting increasing health and nutritional knowledge and awareness, decision-making abilities, access to resources, and behaviour changes (Fadare et al., 2019).

The fourth most important variable of this model, 'Women who currently use any contraceptive method (%)' outlines the relationship between family planning practices and stunting outcomes. Family planning can have an impact on birth spacing and helping families plan their pregnancies and the number of children they can adequately provide for.

'Households with refrigerator (%)' is also one of the 5 most important variables for this model, indicating that households with access to refrigerators were found to have lower rates of stunting. We believe this points out the importance of food safety and food security, through improved food storage and preservation, in child nutrition outcomes, also in line with what was mentioned previously.

The following variables highlighted in this model also reflect the importance of women's empowerment and access to information/media and prenatal care.

3.3.2 Wasting

For wasting outcomes, the top 5 most important variables are related to 'women's education' (wm_liter) and 'power of decision' (wm_decide), 'antenatal care' (wm_4visits, wm_skill), and 'seasonal water variability' (sev).

Women's autonomy in decision-making (wm_decide) is associated with better wasting outcomes in children, possibly due to improved access to health care services and power to make informed decisions that prioritize their own health and the health of their children. Furthermore, having antenatal care with skilled assistance (wm_skill) and adequate follow-up (wm_4visits) may lead to overall improved maternal and childcare, due to a closer monitoring and reduced complications during and after childbirth, which can impact undernutrition outcomes.

3.3.3 Underweight

For the underweight model, the top 5 most important variables again highlight the relevance of 'water supply stability' (sev), 'family planning' (wm_fpuse), 'women's education' (wm_liter), 'decision-making autonomy' (wm_decide), and 'percentage of grassland' (grass_land).

More suitable environmental conditions, including adequate rainfall and fertile soils, giving grassland, lichen mosses or sparse vegetation areas (grass_per), promote practices related to food production and preservation which may contribute to healthier dietary patterns and consequently, better nutritional outcomes for children.

In all three models, the variables related to women's empowerment and possessions appear as highly relevant, such as 'literacy' (wm_liter), 'decision-making' (wm_decide), 'owning a bank account'

(wm_bank), 'contraceptive use' (wm_fpuse) and 'access to no media' at least once a week (wm_nomedia). Among the environmental variables, the 'seasonal water availability variation' (sev) and the' percentage of grassland' (grass_per) are ranked the most important. Besides these, 'antenatal care with skilled assistance' (wm_skill), 'absence of basic vaccination in children' (ch_no_vac) and 'being women in a marriage or union' (wm_union) are also among the most important variables, regardless of the model.

These results indicate that there are common conditions affecting all three undernutrition outcomes and that similar strategies may be suitable to tackle the issues. On the other hand, the results also showed that each undernutrition outcome is also influenced by specific conditions that may require tailored approaches. **Stunting** is specifically influenced by the degree of 'human intervention' (hmts) and 'population density' (pop_dens). **Underweight** shares common drivers with other outcomes, except for three variables: 'nutrition-related variables for women' (wm_thin) and 'reproductive health of mothers' (wm_iron, wm_birth2). **Wasting**, on the other hand, exhibits more distinct aspects, with several exclusive variables:

i. Child-related variables, such as child order 3 or more (ch_order3), lack of vaccination (ch_no_vac), anaemia (ch_anaemia);
ii. Women related variables, such as being married or in a union (wm_union), having employment (wm_employ), using a bank account (wm_bank), use of contraception (wm_fpuse), media exposure (wm_media3, hh_radio, wm_nomedia) and having the first childbirth before being 18 years old (wm_birth18);
iii. Household aspects, such as owning livestock (hh_livestock), use of different types of cooking fuel (hh_solid_fuel, clean fuel), limited access to hand wash (hh_limited), having agricultural land (hh_agriland), having a refrigerator (hh_fridge), and people in the first or second lowest wealth quintiles (hh_poor). Lastly, some environmental variables are especially important to this outcome, which are land productivity (agr_prod), percentage of arid areas (arid_per) and road density (road_dens).

The results obtained for the wasting model indicate that this undernutrition outcome is more complex to understand, having a larger and more diversified set of drivers in comparison with the other models.

3.3.4 Interpreting the combination

To further interpret the results, a brief evaluation was conducted by counting the number of occurrences of variables across all models. When focusing on the variables present in at least two of the three models, the table below is obtained (Table3.8.):

Table 3.8. List of variables that appeared in at least one of the three models tested (Stunting, Wasting, Underweight). 'COUNT' shows the number of models where each variable was found relevant.

VARIABLE	DESCRIPTION	COUNT
wm_liter	Women literacy - higher than secondary or can read part or whole sentence (%)	3
sev	Seasonal water supply variability (index)	3
wm_decide	Women who decide on health, purchases and visits, either alone or jointly with partner (%)	3
wm_fpuse	Women who currently use any contraceptive method (%)	3
grass_per	Grassland, lichen mosses or sparse vegetation (%)	3

wm_skill	Mothers with antenatal care with skilled assistant (%)	3
wm_mobile	Women who own a mobile phone, either alone or jointly (%)	2
wm_4visits	Mothers with antenatal care follow-up with >=4 visits coverage (%)	2
wm_birth18	Women with first birth < 18 years old (%)	2
hh_poor	People in the first or second wealth quintile - poor (%)	2
wm_bank	Women who use a bank account (%)	3
ch_no_vac	Living children aged 12-23 months with no vaccine received (%)	3
wm_house	Women who own a house, either alone or jointly (%)	2
wm_union	Women who are married or in a union (%)	2
wm_nomedia	Women who access none of the three media (tv, radio, newspaper) at least once a week (%)	2

The above results allow for the interpretation of fluctuations in variable rankings across each model and globally. For instance, the variable 'Women who own a mobile phone, either alone or jointly' (wm_mobile), despite being the main driver for the stunting model (Table 3.**8**), is no longer predominant when we analyse its presence across the three models. In fact, people can argue about its added value claiming that it is more an indicator of wealth, as found by Alexandre Abreu (Abreu, 2012), and that wealth itself is likely to be the actual driver, providing a greater capability and awareness for acquiring and dispensing nutritionally improved meals, instead of a mobile phone ownership *per se*.

A similar reasoning could be applied for the variable 'Women literacy - higher than secondary or can read part or whole sentence (%)' (wm_liter). Nevertheless, as it appears predominant across all the three models it clearly demonstrates that literacy contributes to improved nutrition by itself. Literacy is highly correlated with the capability to provide adequate nutrition to one's children – there are more financial resources to pay for the education, the women are freer to attend school, and they have better overall skills. In this way, this is similar to the previous indicator, with the strong benefit of being top of the scale across the models. Moreover, it is relevant to consider the beneficial impact of improved knowledge regarding meal preparation and food diversity. Across the Sahel region, the staple foods are mostly rice, maize, and cassava (Maur & Shepherd, 2015), which are increasingly imported and industrially processed (Haggblade et al., 2017; Reardon, et al., 2021). This leads to a significant reduction in nutritional value, that must be compensated via complementary food elements, which may go against the local culture. For all these reasons, education emerges as important tool to fight undernutrition.

The 'seasonal water supply variability (index)' (sev) indicator, which is directly connected to 'grass_per', offers a distinct perspective on the challenges faced by women in ensuring the nourishment of their children. This indicator tops the list – it ranks #2 in the stunting model, #5 in wasting, and #1 in underweight – and is on a par with 'wm_liter'. This challenge is not directly correlated with economic, social, or governmental failures, commonly observed across the Sahel region; instead, it is often influenced by a complex interplay of potentially conflicting factors. The primary factor contributing to this challenge is the depletion of the water table, resulting in increasingly scarce water availability during the dry season. Consequently, not only women and children have to travel greater distances, sometimes up to four kilometres, to access drinking water by the end of the dry season, but also there is poorer water quality – often becoming mixed with mud. A secondary factor is the time and energy expended to obtain water,

which limits the time available for managing even basic subsistence farming activities. This, in turn, leads to a substantial reduction in the diversity of available nutrients.

4 Conclusions

This study investigated the associations between child, women/mother, household, and environmental-related variables with stunting, wasting, and underweight outcomes in children under five, in the Sahel region, using ML methods.

Several steps were undertaken to collect, process, and analyse the available data. Emphasis was placed on reliable and representative data sources, such as the Demographic and Health Surveys (DHS) for population data and other publicly available datasets for environmental variables. Ensuring data accuracy and spatial precision was crucial for subsequent analysis and elaborating conclusions.

The pre-processing steps were aimed at maintaining spatial accuracy, respecting administrative levels, and performing exploratory statistical analyses. This process was critical for drawing insights, identifying patterns, and uncovering relationships within the data, enabling the selection of the appropriate approach, and ensuring optimal performance of the machine learning methods. The statistical analysis results were based on dataset with selecting variables that accurately represented the regional aspects relevant to undernutrition. Overall, the methodological protocol design lends robustness.

The chosen models were then applied, and their performance metrics were evaluated. RF was chosen as the best method for modelling stunting, wasting, and underweight drivers, demonstrating slightly lower performance for wasting, but excellent performance for stunting, and an overall better performance than BRT.

Variables related to women's empowerment, such as decision-making autonomy, literacy, and ownership of mobile phones, as well as variables related to healthcare access, such as vigilance during pregnancy and family planning practices, emerged as important predictors of improved nutritional outcomes in children under five, in the Sahel region. Environmental factors, such as variations in water supply and the presence of grassland or vegetation, were also shown to play an important role.

This study was focused specifically on testing ML methods to identify undernutrition drivers, which has provided insightful findings and a partial understanding of this key Public Health issue in the Sahel. By using advanced analytical techniques such as ML to uncover complex relationships and drivers of undernutrition, it was possible to process large volumes of data and identify non-linear patterns and interactions that could have not been apparent through traditional statistical methods. In addition, ML algorithms can handle multidimensional data and account for complex interactions between several socio-economic, health- and nutrition-related and environmental data, having provided a more nuanced understanding of the underlying drivers of undernutrition.

By bringing light to the complex interplay between the multiple undernutrition drivers, targeted interventions and informed policies can be implemented to improve children's overall nutrition outcomes. Nevertheless, further investigation is required. This study is integrated in a more comprehensive framework that includes a systematic literature review and local stakeholder engagement, components which may provide additional insights for a more holistic understanding of the drivers of undernutrition in the Sahel.

5 References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433–459. https://doi.org/10.1002/wics.101.

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A. E., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938. https://doi.org/10.1016/j.heliyon.2018.e00938.

Abreu, A. (2012). Migração e diferenciação de classes na Guiné-Bissau rural. E-cadernos Ces, 15. https://doi.org/10.4000/eces.955.

Adler, E. D., Voors, A. A., Klein, L., Macheret, F., Braun, O. O., Urey, M. A., Zhu, W., Sama, I., Tadel, M., Campagnari, C., Greenberg, B., & Yagil, A. (2020). Improving risk prediction in heart failure using machine learning. *European Journal of Heart Failure*, *22*(1), 139–147. https://doi.org/10.1002/ejhf.1628.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, 1937-1967. https://doi.org/10.1007/s10462-020-09896-5.

Bitew, F. H., Sparks, C. S., & Nyarko, S. H. (2021). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. *Public Health Nutrition*, 25(2), 269–280. https://doi.org/10.1017/S1368980021004262.

Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324.

Capinha, C., Ceia-Hasse, A., Kramer, A. M., & Meijer, C. (2021). Deep learning for supervised classification of temporal data in ecology. Ecological Informatics, 61, 101252. https://doi.org/10.1016/j.ecoinf.2021.101252.

Chilyabanyama, O. N., Chilengi, R., Simuyandi, M., Chisenga, C. C., Chirwa, M., Hamusonde, K., Saroj, R. K., Iqbal, N. T., Ngaruye, I., & Bosomprah, S. (2022). Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia. *Children*, 9(7). https://doi.org/10.3390/children9071082.

Cramer, R. D. (1993). Partial Least Squares (PLS): Its strengths and limitations. *Perspectives in Drug Discovery and Design*, *1*(2), 269–278. https://doi.org/10.1007/BF02174528.

Curto, J. D., & Pinto, J. C. (2011). The corrected VIF (CVIF). *Journal of Applied Statistics*, 38 (7), 1499-1507, DOI: 10.1080/02664763.2010.505956.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Fadare, O., Amare, M., Mavrotas, G., Akerele, D., & Ogunniyi, A. (2019). Mother's nutrition-related knowledge and child nutrition outcomes: Empirical evidence from Nigeria. *PLOS ONE*, 14(2), e0212775. https://doi.org/10.1371/journal.pone.0212775.

Fenta, H. M., Zewotir, T., & Muluneh, E. K. (2021a). A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative

zones. *BMC Medical Informatics and Decision Making*, 21(1), 1–13. https://doi.org/10.1186/s12911-021-01652-1.

Fenta, H. M., Zewotir, T., & Muluneh, E. K. (2021b). Spatial data analysis of malnutrition among children under five years in Ethiopia. *BMC Medical Research Methodology*, 21(1), 1–13. https://doi.org/10.1186/s12874- 021-01391-UNICEF. (2021).

Goldstein, B. A., Navar, A. M., & Carter, R. E. (2017). Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges. *European Heart Journal*, *38*(23), 1805–1814. https://doi.org/10.1093/eurheartj/ehw302.

Haggblade, S., Me-Nsope, N. M., & Staatz, J. M. (2017). Food security implications of staple food substitution in Sahelian West Africa. Food Policy, 71, 27–38. https://doi.org/10.1016/j.foodpol.2017.06.003.

He, Q. P., & Wang, J. (2007). Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, *20*(4), 345–354. https://doi.org/10.1109/TSM.2007.907607.

Hijmans, Robert J., Steven Phillips, John Leathwick, and Jane Elith. 2022. *Dismo: Species Distribution Modeling*. https://rspatial.org/raster/sdm/.

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/rsta.2015.0202.

Knol, M. J., Vandenbroucke, J. P., Scott, P., & Egger, M. (2008). What Do Case-Control Studies Estimate? Survey of Methods and Assumptions in Published Case-Control Research. *American Journal of Epidemiology*, *168*(9), 1073–1081. https://doi.org/10.1093/aje/kwn217.

LeFevre, A., Shah, N., Bashingwa, J. J. H., George, A., & Mohan, D. (2020). Does women's mobile phone ownership matter for health? Evidence from 15 countries. BMJ Global Health, 5(5), e002524. https://doi.org/10.1136/bmjgh-2020-002524.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A Survey of Convolutional Neural Networks: Analysis, applications, and Prospects. IEEE Transactions on Neural Networks and Learning Systems, 33(12), 6999–7019. https://doi.org/10.1109/tnnls.2021.3084827.

Mansur, M., Afiaz, A., & Hossain, Md. S. (2021). Sociodemographic risk factors of under-five stunting in Bangladesh: Assessing the role of interactions using a machine learning method. *PLOS ONE*, *16*(8), 1–17. https://doi.org/10.1371/journal.pone.0256729.

Maur, J. C., & Shepherd, B. (2015). Connecting Food Staples and Input Markets in West Africa: A Regional Trade Agenda for ECOWAS Countries: Report No. 97279-AFR. The International Bank for Reconstruction and Development /The World Bank. https://doi.org/10.13140/RG.2.1.3548.0160.

McKillup, S., & Dyar, M. D. (2010). Geostatistics Explained. An Introductory Guide for Earth Scientists. CAMBRIDGE UNIVERSITY PRESS.

Ogunniyi, A., Omotoso, S. O., Salman, K. K., Omotayo, A. O., Olagunju, K. O., & Aremu, A. O. (2021). Socio-economic Drivers of Food Security among Rural Households in Nigeria: Evidence from Smallholder Maize Farmers. *Social Indicators Research*, 155(2), 583–599. https://doi.org/10.1007/s11205-020-02590-7.

Parmar, A., Katariya, R., & Patel, V. (2018). A review on Random Forest: An ensemble classifier. In Lecture notes on data engineering and communications technologies (pp. 758–763). https://doi.org/10.1007/978-3-030-03146-6_86.

Reardon, T., Tschirley, D. L., Liverpool-Tasie, L. S. O., Awokuse, T. O., Fanzo, J., Minten, B., Vos, R., Dolislager, M., Sauer, C., Dhar, R., Vargas, C. M., Lartey, A., Raza, A., & Popkin, B. M. (2021). The processed food revolution in African food systems and the double burden of malnutrition. Global Food Security, 28, 100466. https://doi.org/10.1016/j.gfs.2020.100466.

Saroj, R. K., Yadav, P. K., Singh, R., & Chilyabanyama, O. N. (2022). Machine Learning Algorithms for understanding the determinants of under-five Mortality. *BioData Mining*, 15(1), 1–22. https://doi.org/10.1186/s13040-022-00308-8.

UNICEF (2021). UNICEF Conceptual Framework on Maternal and Child Nutrition. UNICEF. https://www.unicef.org/media/113291/file/UNICEF%20Conceptual%20Framework.pdf.

Wei, W., Ramalho, O., Malingre, L., Sivanantham, S., Little, J. C., & Mandin, C. (2019). Machine learning and statistical models for predicting indoor air quality. *Indoor Air*, *29*(5), 704–726. https://doi.org/10.1111/ina.12580.

World Health Organization (WHO). (2024, 01/Mar). Malnutrition. https://www.who.int/news-room/fact-sheets/detail/malnutrition.

Woldesenbet, B., Alemu, T., & Tsegaye, B. (2023). Water, hygiene and sanitation practices are associated with stunting among children of age 24-59 months in Lemo district, South Ethiopia, in 2021: community based cross sectional study. *BMC Nutrition*, *9*(1). https://doi.org/10.1186/s40795-023-00677-1.

Young, H. (2020) Nutrition in Africa's drylands: Aconceptual framework for addressing acute malnutrition. Boston: Feinstein International Center, Tufts University, 2020.

6 Appendixes

Appendix I – Detailed description on the methodological approach

6.1 Geodatabase

To investigate the drivers for undernutrition in the Sahel region using machine learning methods and AI modelling, a set of indicators and variables was identified in accordance with the specific literature and the UNICEF framework. The Demographic and Health Survey (DHS) database was the main source of data on the undernutrition outcomes and on the potential direct and immediate drivers of undernutrition, regarding child and mother characteristics, household assets and birth conditions, among others. For the systemic (basic) drivers, other databases with freely available high-resolution data sources were searched, covering environmental and social conditions.

The dependent variables considered to better translate the prevalence of undernutrition in the Sahel region, according to the literature, were stunting, wasting and underweight in children under the age of five. As described in the previous report, there were some aspects of the variables that needed to be accounted for, such as the classification discrepancies among samples for some indicators (i.e. religion, ethnicity), the absence of data for a particular country, and inconsistent temporal resolution of the data among countries. Moreover, the need to accommodate variables from different sources required a spatial transformation process of aggregation of the individual or household data by sampling units (DHS clusters), followed by an estimation (mean value of each variable) by the sub-regional level (administrative units level 2 - adm2). This aggregation of data by the administrative regions guaranteed that the estimations were unaffected by the clusters' displacement due to privacy protection of the interviewees. It should also be noted that, because the DHS Program uses both oversampling and undersampling for stratification purposes, there was the need to recalculate the variables for each cluster, using sampling weights, to maintain statistical accuracy, before proceeding with the estimations for the administrative regions (sub-regional level, adm2).

Due to the complexity of the database, both the extraction of the raw data and the estimation of the indicators from DHS were obtained by applying automated routines in R language, based on a toolset provided by the DHS Program's Code Share. The indicators can be retrieved in both tabular (excel) and vectorial (shapefile) formats, the latter required to create a geodatabase which can be incorporated into a geographic information system, or other spatial data analysis platform.

The first version of the geodatabase was part of Deliverable 1 of this project. After the second progress meeting, held on the 23rd of February 2023, and following the updated results of the literature review, it was agreed that an additional search for other potential indicators would be carried out. A revised version of the Geodatabase was delivered later on, together with an additional technical report describing in detail the process of calculating the indicators from DHS data for the sub-regional level, for all the Sahel countries considered (available upon request).

6.2 Data Pre-Processing

Pre-processing of raw data is crucial to obtain a reliable data set and apply statistical analysis and machine-learning algorithms. This process implied dealing with inconsistent/incomplete data. After the conversion of raw data into a set of variables, the ones with missing values for one or more administrative areas were not included in the geodatabase, since performing spatial analysis with inconsistent spatial distribution

leads to inaccurate results. Additionally, all the DHS clusters that did not provide spatial reference were excluded from the study.

Data harmonization was performed through the conversion of all the variables into the same format and spatial resolution (normalization), thus reducing the unwanted influence of variations between samples, and ensuring statistical and spatial comparisons. This implied transforming discrete variables into continuous data, by the conversion of absolute values of a given category into a percentage, where applicable.

While machine-learning methods do not require normal distributions of data, statistical exploratory methods do (e.g. clustering); therefore, standardization of the data was performed, by calculating the z-scores for each variable. This measurement indicates how far from the mean a data point is, in terms of standard deviations, allowing for comparison between the data and a normally distributed population.

Finally, the existing relationships between variables were explored, mainly to verify the presence of collinearity. Strong correlations between explanatory variables are problematic for machine-learning algorithms because it interferes with the model's capacity to determine how much each variable, independently, affects the outcomes. This reduces the model's precision, weakening its statistical and explicative power. The collinearity issues were assessed by applying the correlation coefficient and the variance inflation factor (VIF). The first is a bivariate measurement of the strength of association between pairs of variables, while the second evaluates how much the variance of an independent variable is inflated by its correlation with the other independent variables. These have proven to be two robust and efficient techniques to identify the data that might need to be excluded from the modelling procedure, to avoid duplicate effects in the response variable.

6.3 Exploratory Statistical Methods

Exploratory analysis was used to scout the data structure and investigate trends in the data set. For this purpose, three linear statistical multivariate methods were applied - principal component analysis (PCA), partial least squares (PLS) and k-nearest neighbours (kNN).

6.3.1 Principal Components Analysis (PCA)

A consolidated unsupervised exploratory data analysis tool that reduces the dimensionality and complexity of the data set, thus enabling a better understanding of its structure and inner relationships (Abdi & Williams, 2010; Jollife & Cadima, 2016). This method works by combining highly correlated variables, producing a reduced set of new uncorrelated ones that best describe the variability in the data. These new variables, called principal components, are listed in decreasing order of importance. The result also shows the contribution of the original variables for each component (Abdi & Williams, 2010; McKillup & Dyar, 2010).

6.3.2 Partial Least Squares (PLS)

A linear regression model that estimates the relationships among variables, computing changes in the output according to changes in the input, based on the system properties (Cramer, 1993). Moreover, PLS projects input and output variables into a new space, obtaining variance-covariance matrices, resulting in new components to better explain the output variables. Therefore, it reduces the dimension of the original variables, while maintaining the most relevant information (Wei et al., 2019).

6.3.3 kNN

A non-parametric, supervised-learning classifier that examines multidimensional distances to compute groupings of data. More precisely, it determines the nearest neighbours of a given point, using a distance metric, the most common being the Euclidian distance. The distances are calculated between the new sample and a set of previous measurements (training set); then, it assigns a label (class) based on two main voting schemes - majority voting and weighted-sum voting. The first assigns the class that appears most frequently within the k-nearest neighbours, while the second assigns the class based on the idea that proximity indicates similarity, attributing weights according to this assumption (Wei et al., 2019). The main advantage of kNN is its ability to describe complex nonlinear boundaries between groups of data and it is a valuable method to map spatial patterns of the variables and identify their (dis)similarities; nevertheless, it does not deliver information about significant compounds (He & Wang, 2007), nor provides meaningful explanatory outcomes of the patterns found.

6.4 Machine Learning Methods and AI Modelling

ML approach assumes that the data-generating process is complex and unknown, and instead of applying a statistical model to the data, it uses an algorithm to learn the relationships between outputs and inputs and predict future responses, based on observed dominant patterns. Therefore, these methods effectively investigate nonlinear statistical relations between variables, which may stand out in the presence of contrasting subgroups of data.

The ML methods applied were algorithms based on decision trees (Random Forest (RF) and Boosted Regression Trees (BRT)), and on Artificial Neural Networks (ANN). These were considered the most straightforward and robust to address the complexity of our data and obtain outputs able to explain undernutrition patterns. The questionnaires of DHS were carried out in distinct years for the different countries, and it was not possible to collect data for more than one timeframe for each Sahel country. For this reason, it was not possible to apply algorithms based on Convolutional Neural Networks (CNN), because these require time-series data to work properly. A feedforward ANN algorithm was applied instead.

6.4.1 Random Forest (RF)

RF is a non-parametric, ensemble ML model that builds multiple uncorrelated decision trees (classifiers), each one based on a randomly selected subset of the original samples, outputting the mean prediction of the individual trees. This advanced ensemble tree approach is based on solid statistical theories, hence outperforming other methods in terms of classification accuracy (Breiman, 2001; Cutler et al., 2007; Elith et al., 2008; Wei et al., 2019).

RF has very good accuracy, high flexibility to compute several types of statistical data analysis, an ability to deal with missing values, outliers and mislabelled samples (Cutler et al., 2007) and it also overcomes the overfitting problem (Breiman, 2021; Wei et al., 2019). The challenges are related to parameter selection, interpretation of results and computation time. The error rate of the RF model depends on two things: the correlation between trees and the strength of the individual trees (Elith et al., 2008). Given its strengths and the fact that a single regression tree algorithm may be less accurate to model complex problems (Wei et al., 2019), RF was considered one of the most robust methods for dealing with the complex interactions among the variables contemplated in this study.

6.4.2 Boosted Regression Trees (BRT)

The BRT is a self-learning decision tree-based algorithm that combines a large number of tree models to optimize predictive performance. Instead of fitting a large decision tree to the data, with potential overfitting, the boosting approach fits the model iteratively (and more slowly) to the training data. The first regression tree reduces the loss function, while the second one is fitted to the residuals. The model is then updated, and the residuals are recalculated; this step is repeated until the criterion is met (e.g. a specific number of iterations or minimal residuals). Boosting process is unique because it is stagewise, meaning that the existing trees are unaffected by the expansion of the model; it also means that the construction of each tree depends on the trees that have already been created. The final BRT model translates the sum of all trees multiplied by the learning rate (Elith et al., 2008). This model is very similar to RF, as both build many uncorrelated classifiers, however, while RF focuses on randomization, BRT focus on misclassified samples of previous models.

6.4.3 Artificial Neural Networks (ANN)

ANNs are self-learning algorithms designed to mimic the human brain and perform data classification and prediction through pattern recognition (Abiodun et al., 2018; Wei et al., 2019). Their architecture is defined by an interconnected structure of nodes (neurons) organized in layers (input layer, hidden layer/s and output layer), where each node is connected to neurons in the previous layer, similar to human neuron processing units. These algorithms use a combination of weights and functions to compute estimations (outputs) from input variables, employing backpropagation for training the algorithm. ANNs can operate complex systems with large inputs of data and deal effectively with non-linear problems. Other strengths are their strong fitting ability, self-learning, fault tolerance, and advancement in input to an output mapping. On the other hand, the training of the system is computationally intensive and time-consuming (Abiodun et al., 2018; Wei et al., 2019).

Appendix II - Mapping Results

This section presents the mapping outputs of the modelling procedure applied, with the spatial distribution of the drivers (values of variables) selected per model.

The classes defined for mapping the variables were based on the most suitable classification scheme related to the available range of values and to facilitate visualization, either through quintiles, natural breaks, or equal breaks.

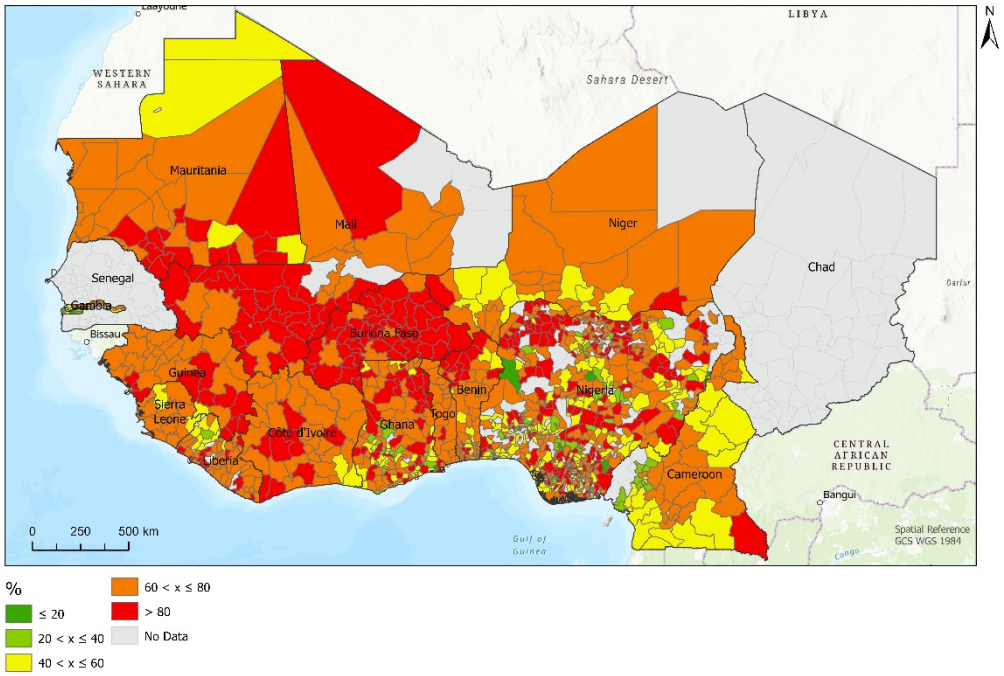

Figure II.1. Child prevalence of anaemia (%).

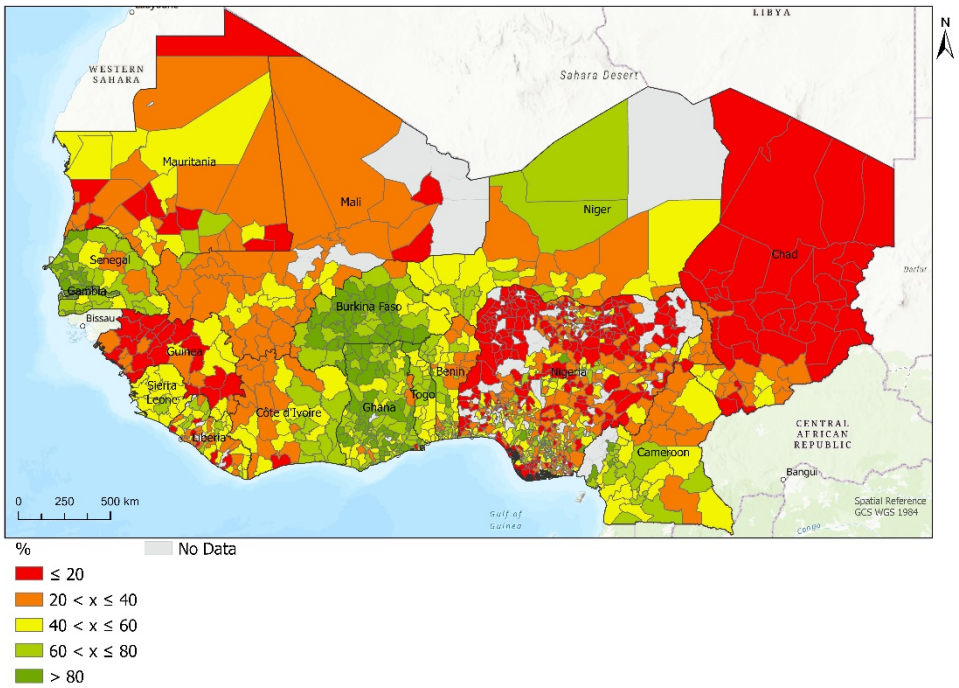

Figure II.2. Living children aged 12-23 months who received all basic vaccination (%).

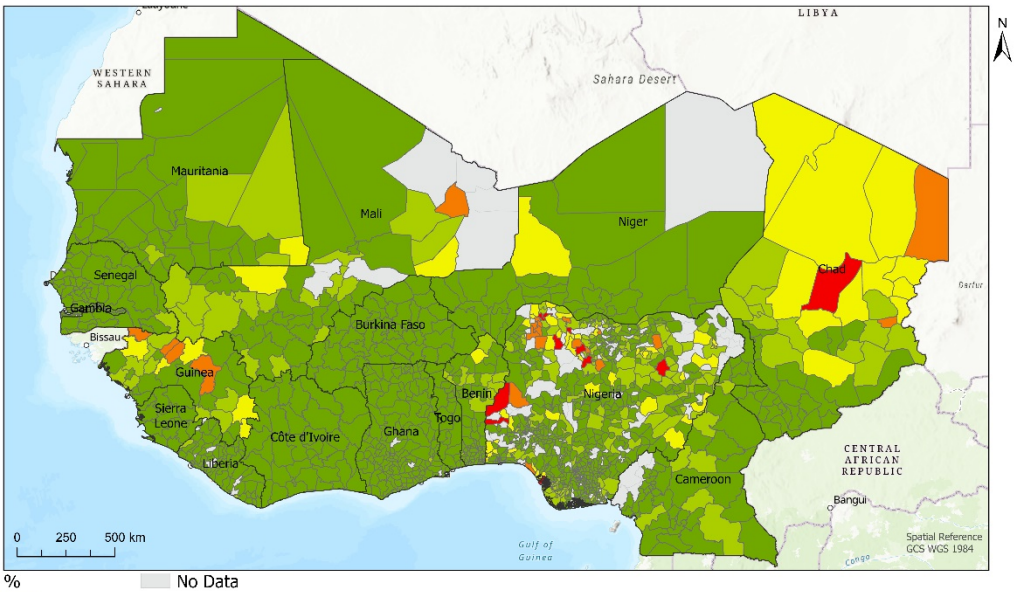

Figure II.3. Living children aged 12-23 months with no vaccine received (%).

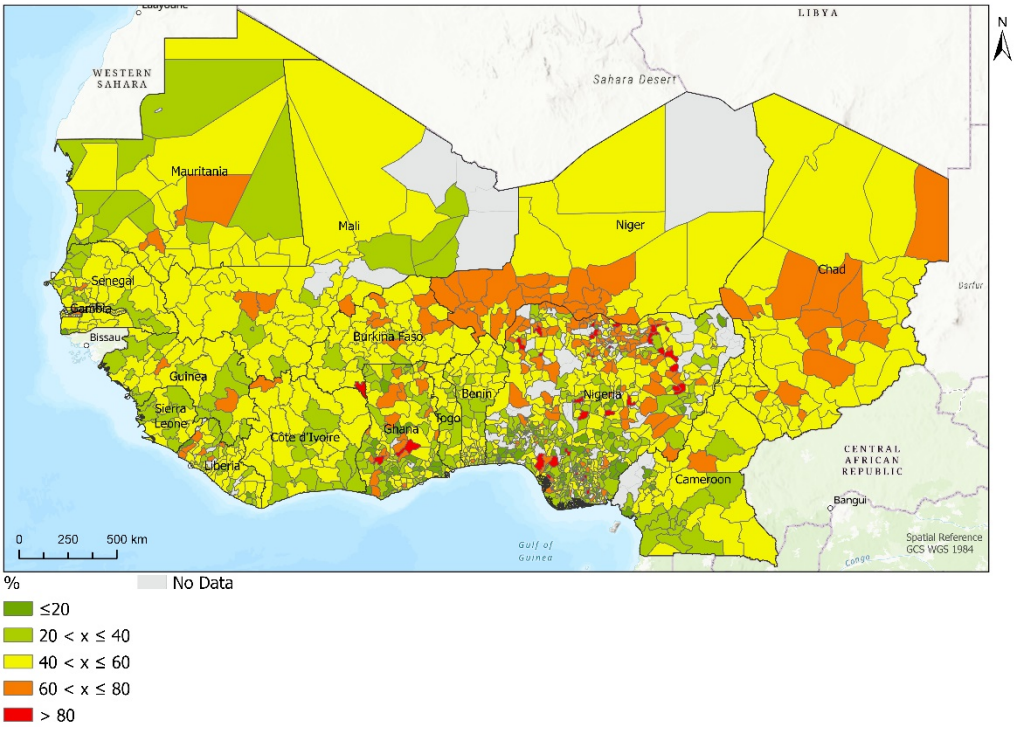

Figure II. 4. Children 3rd order or more (%).

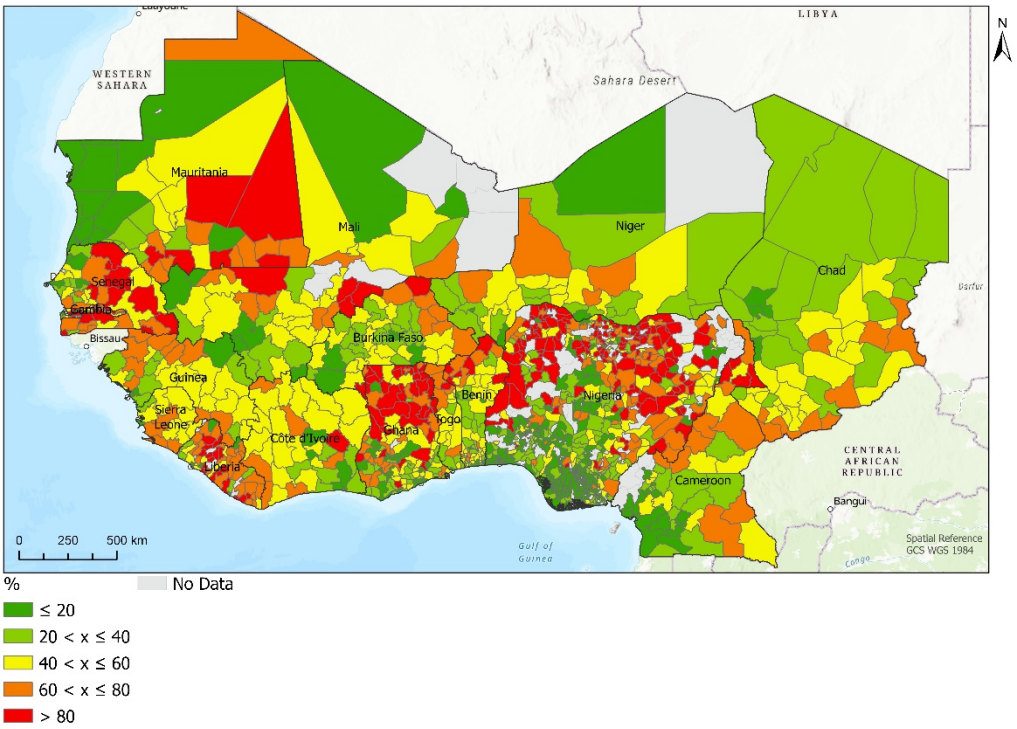

Figure II.5. People in the first or second wealth quintile (poor) (%).

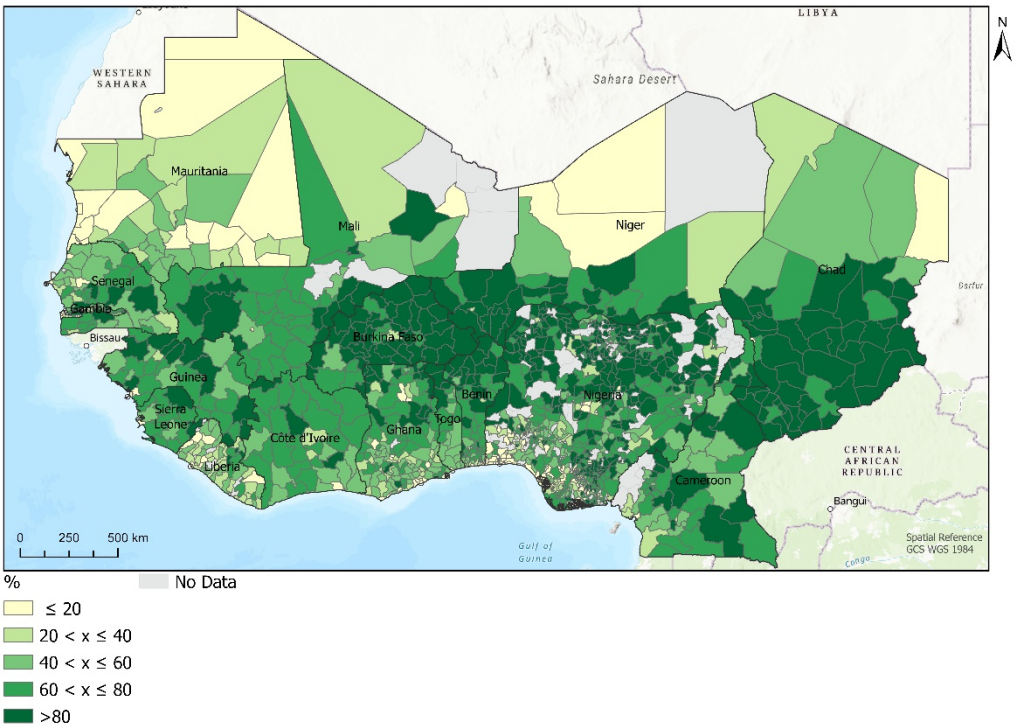

Figure II.6. Households with agricultural land (%).

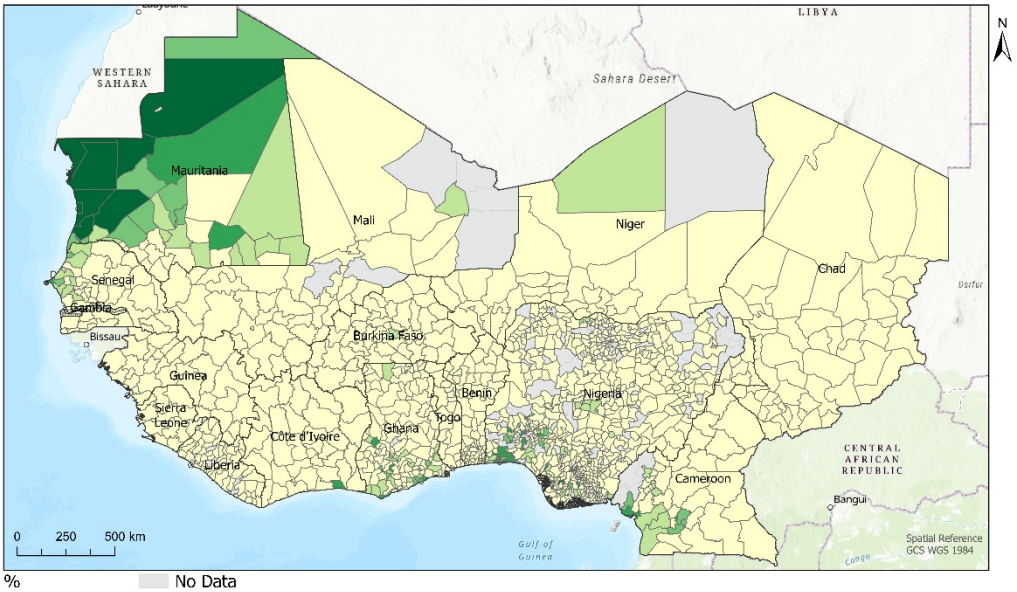

Figure II.7. Households using clean fuel for cooking (%).

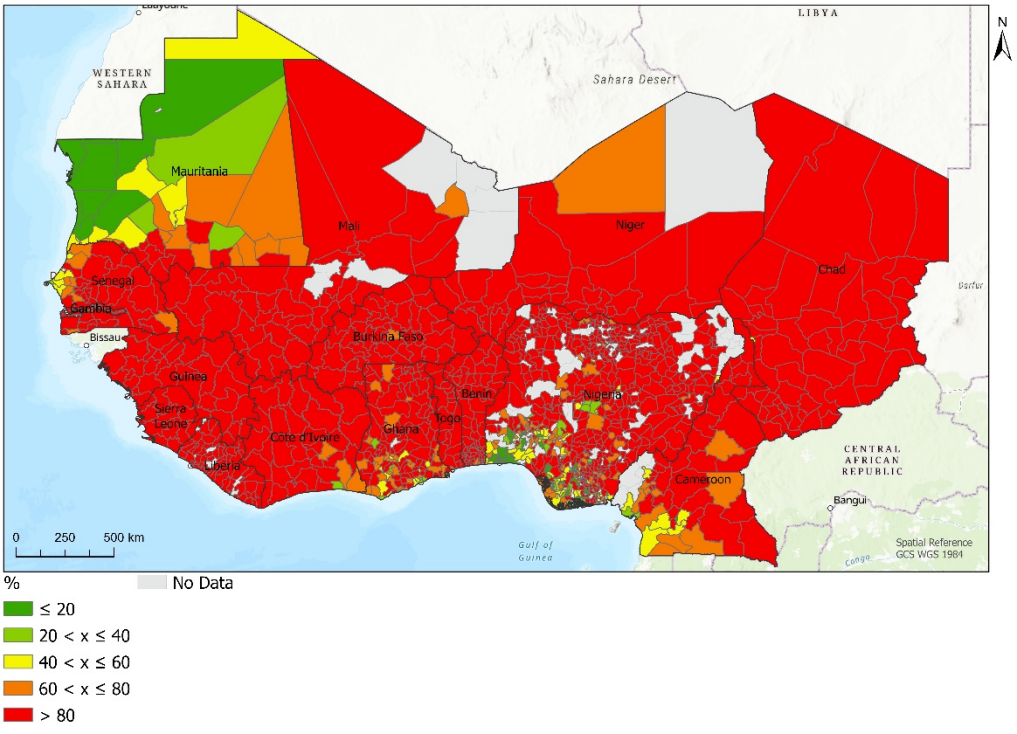

Figure II.8. Households using solid fuel for cooking (%).

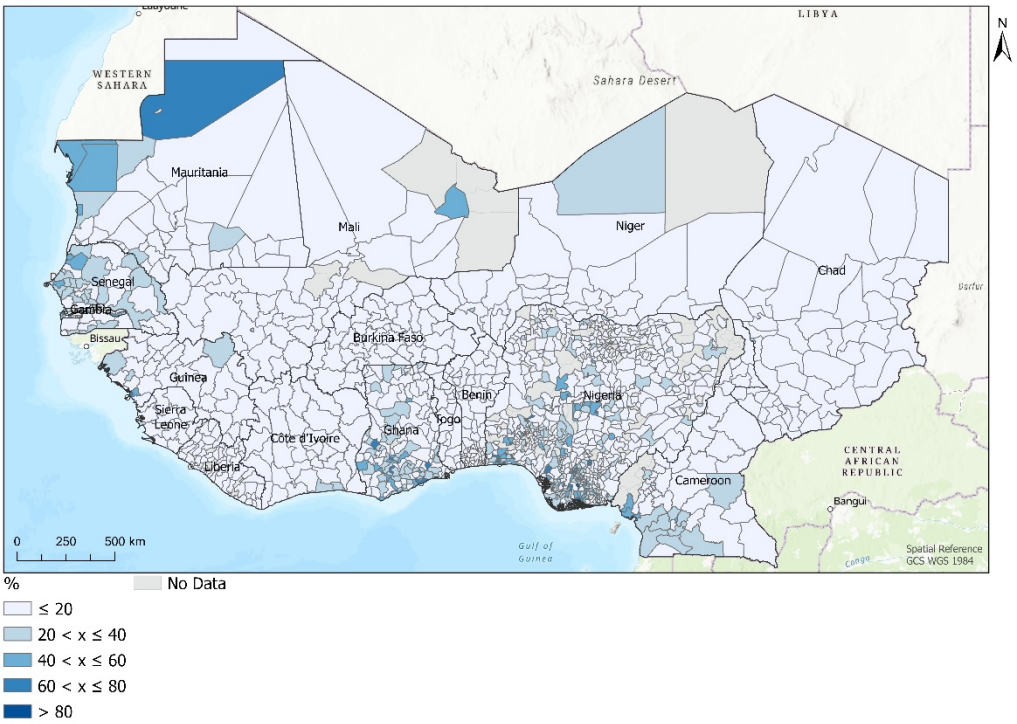

Figure II.9. Households with refrigerator (%).

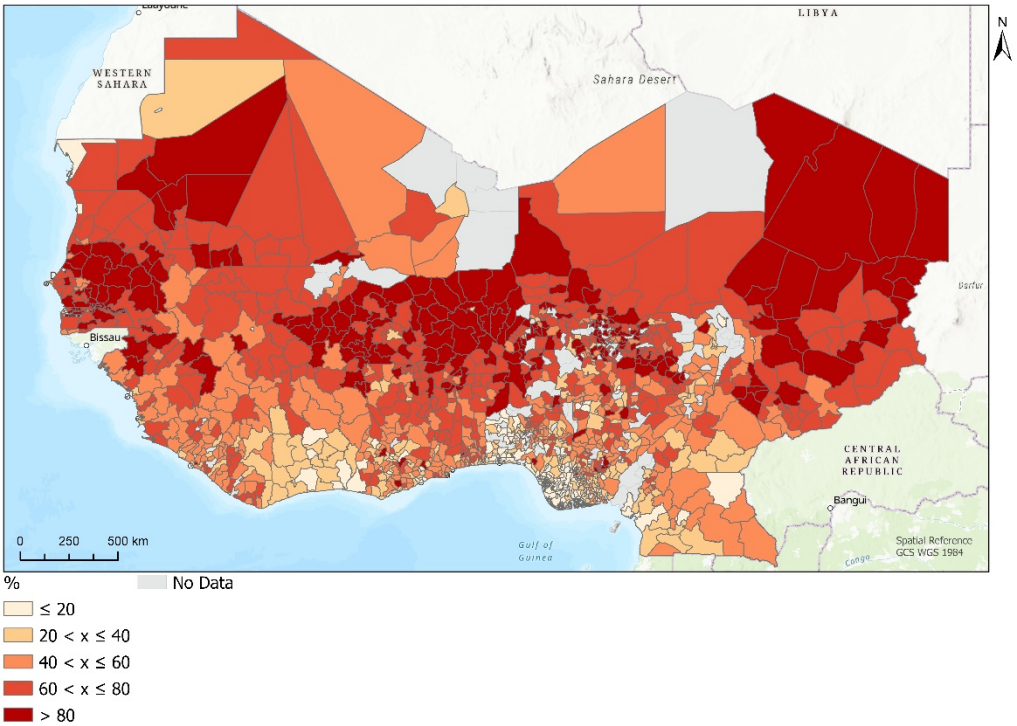

Figure II.10. Households with livestock (%).

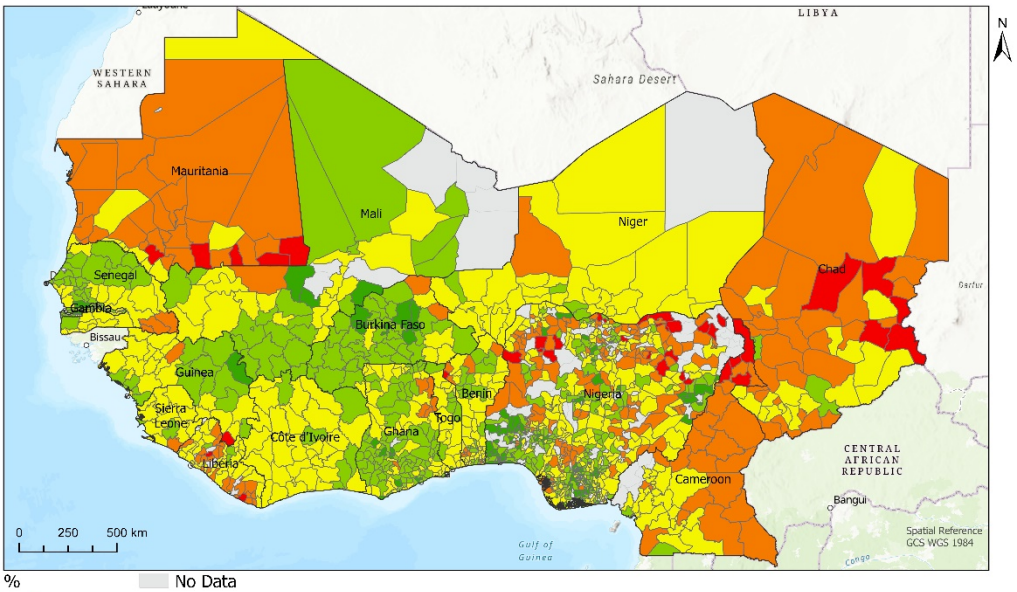

Figure II.11. Households with radio (%).

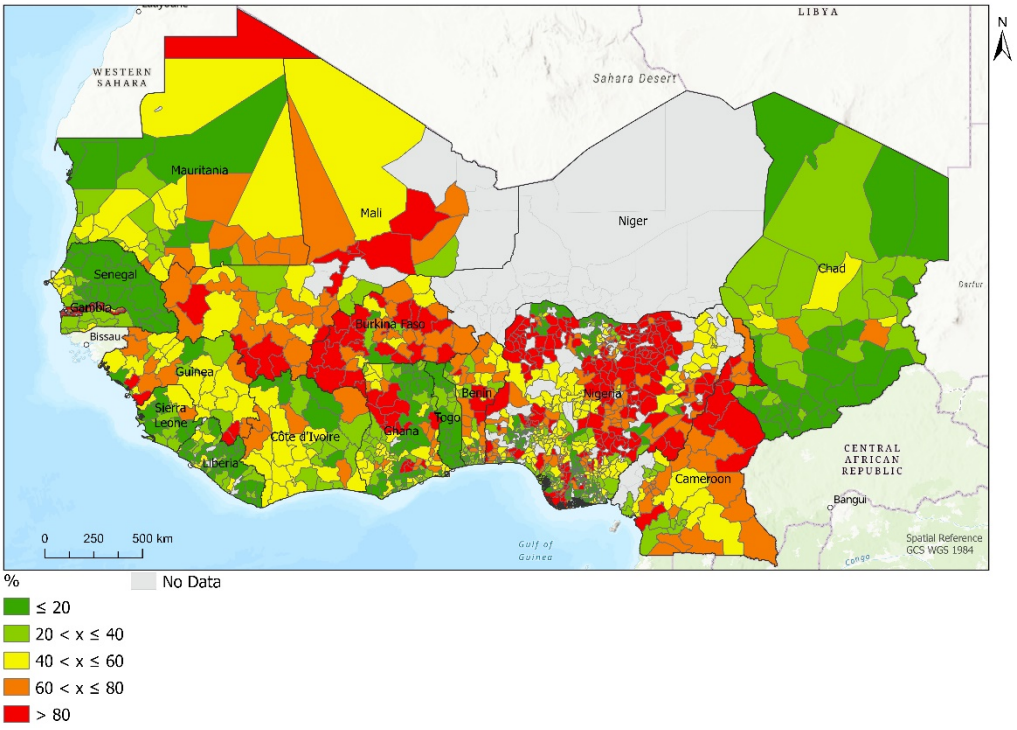

Figure II.12. Households with limited hand washing facilities (%).

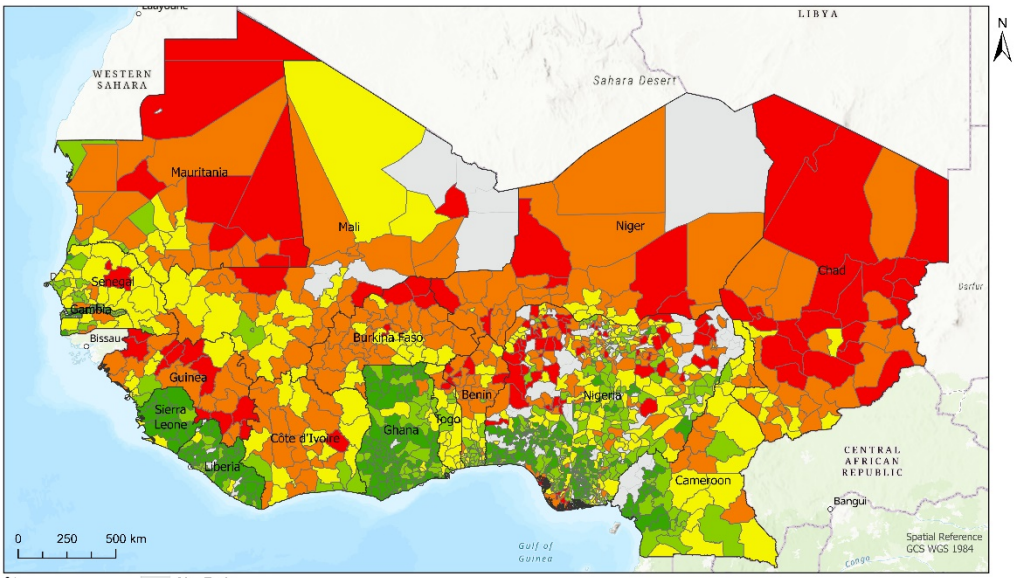

Figure II.13. Mothers with antenatal care follow-up with ≥4 visits coverage (%).

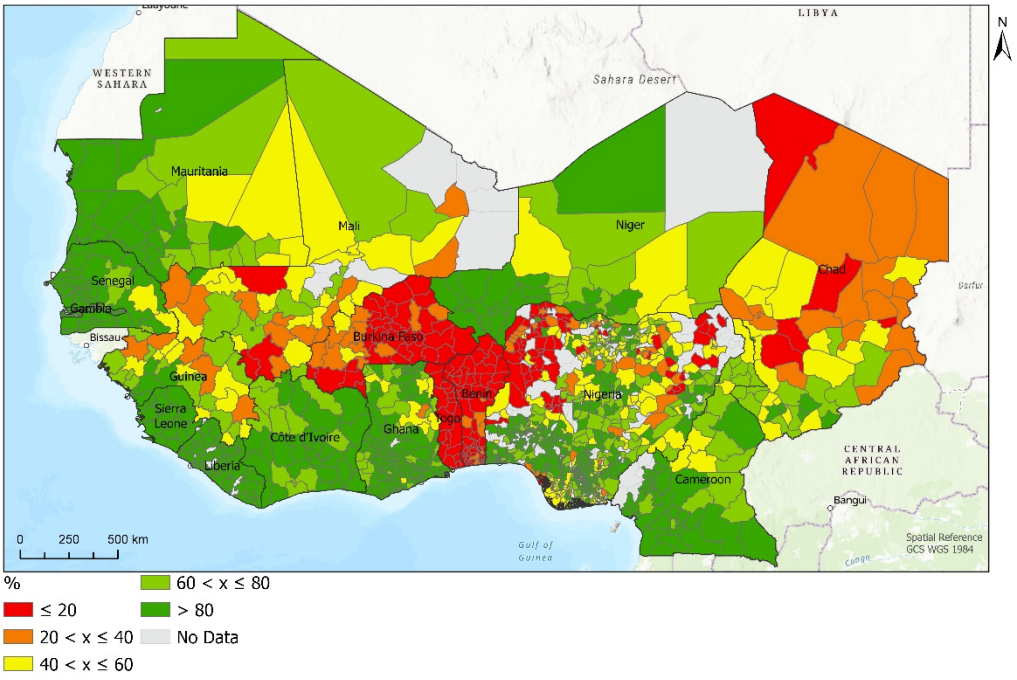

Figure II.14. Mothers with antenatal care with skilled assistance (%).

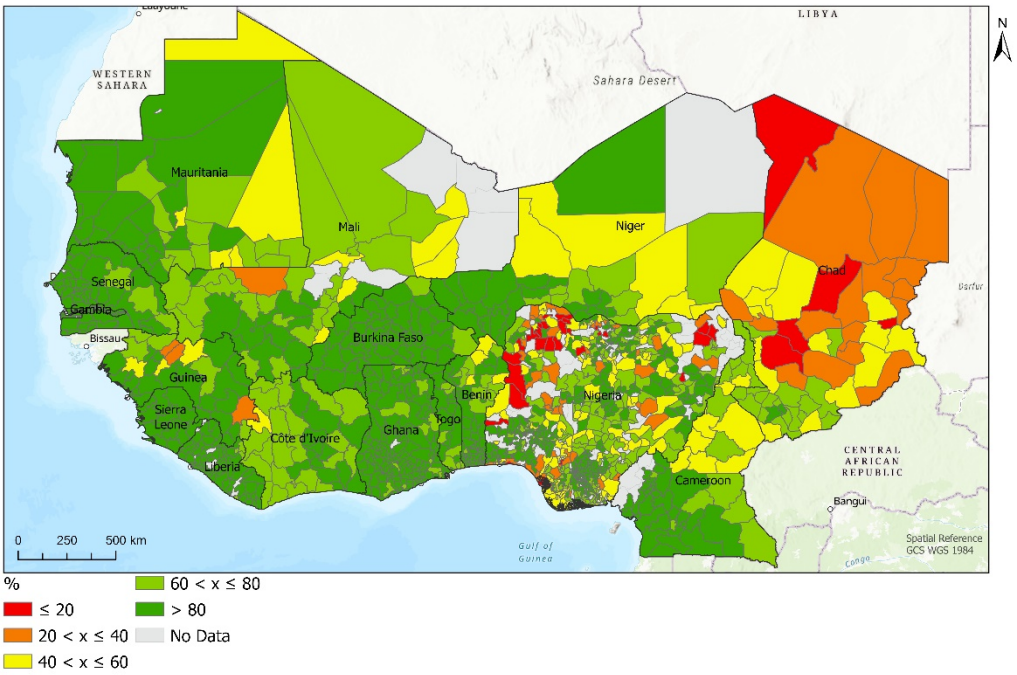

Figure II.15. Mothers with iron supplementation during pregnancy (%).

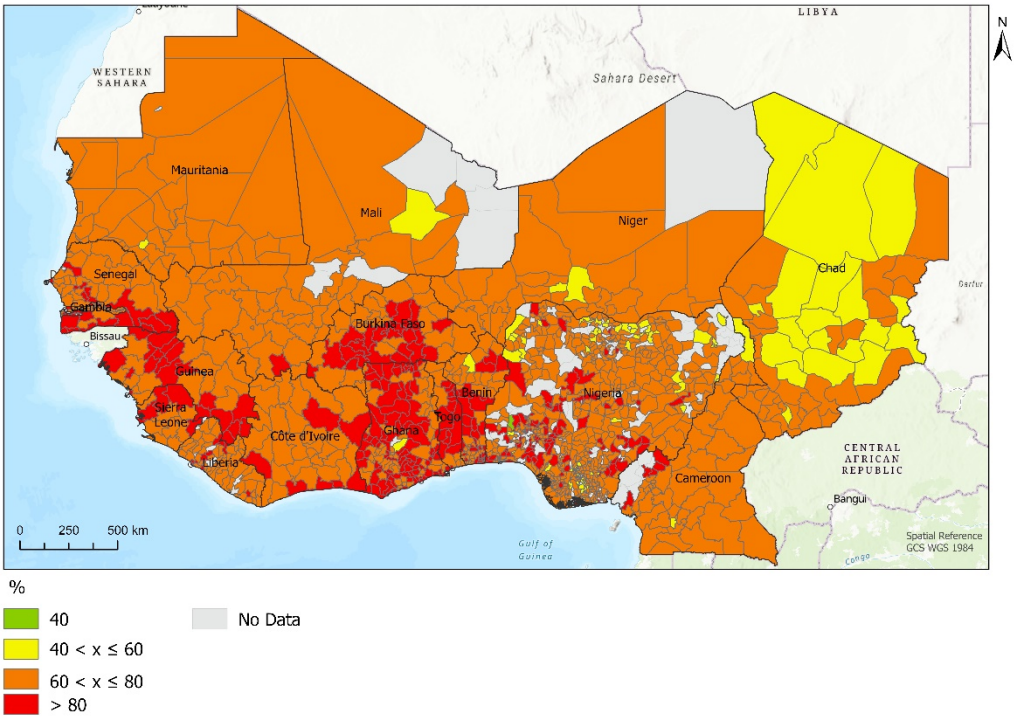

Figure II.16. Women with birth interval < 2 years (%).

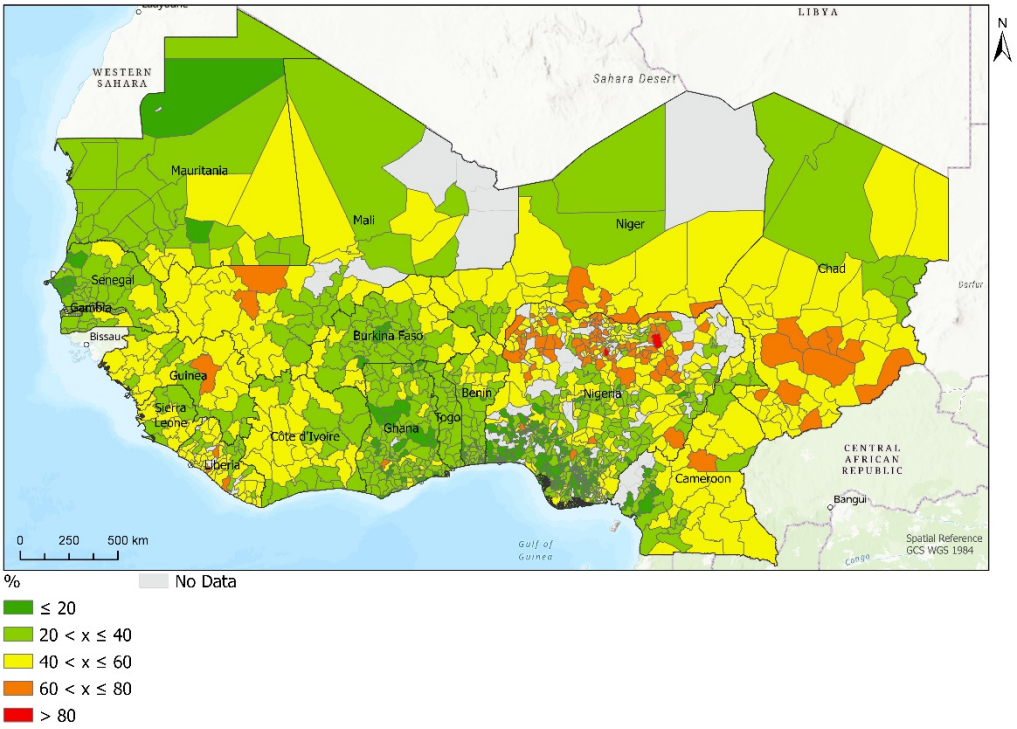

Figure II.17. Women with first birth < 18 years old (%).

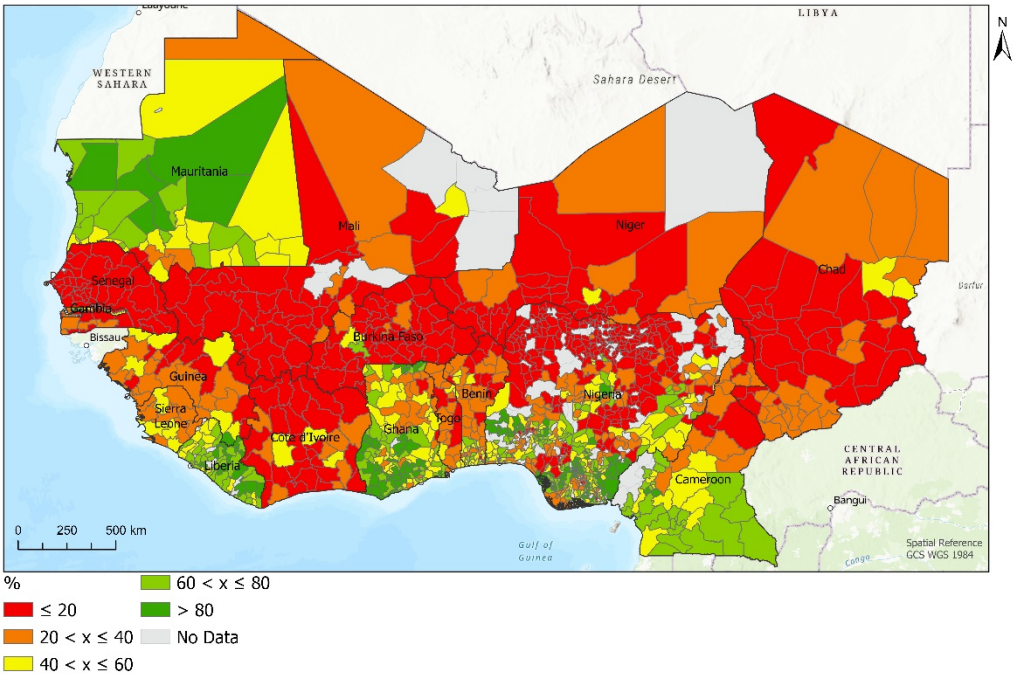

Figure II.18. Women who decide on health, purchases and visits, either alone or jointly with a partner (%).

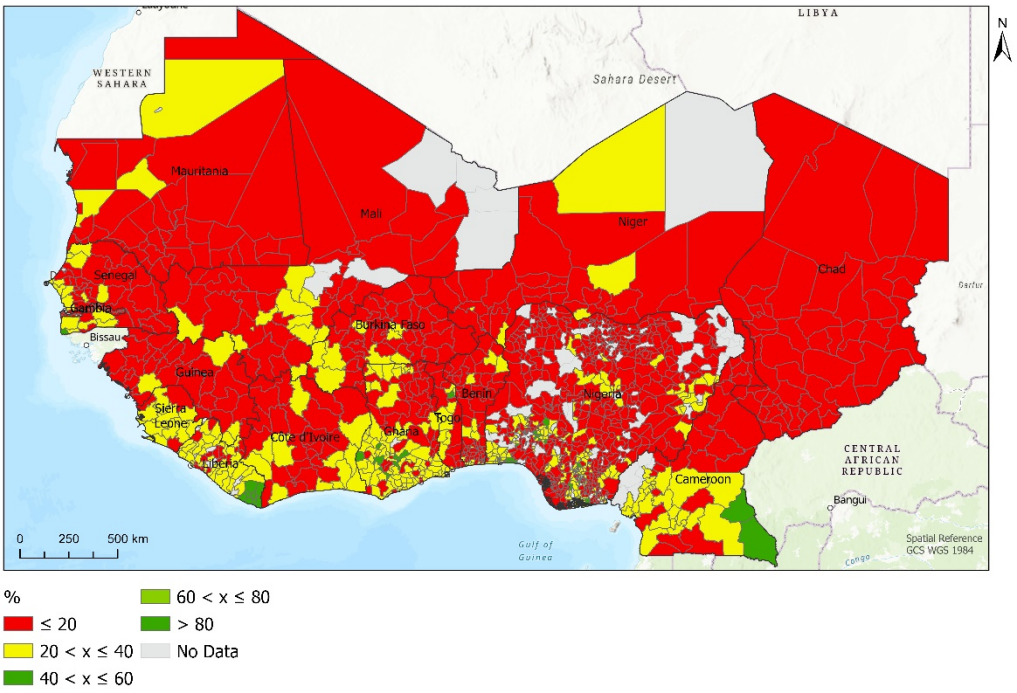

Figure II.19. Women who currently use any contraceptive method (%).

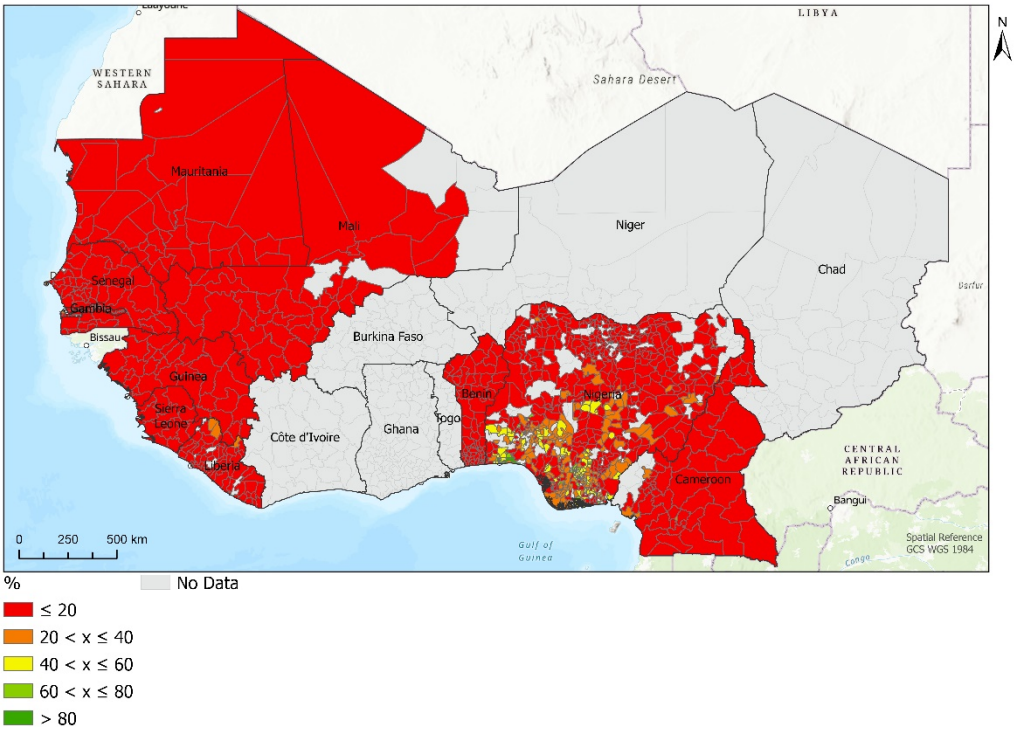

Figure II.20. Women who use a bank account (%).

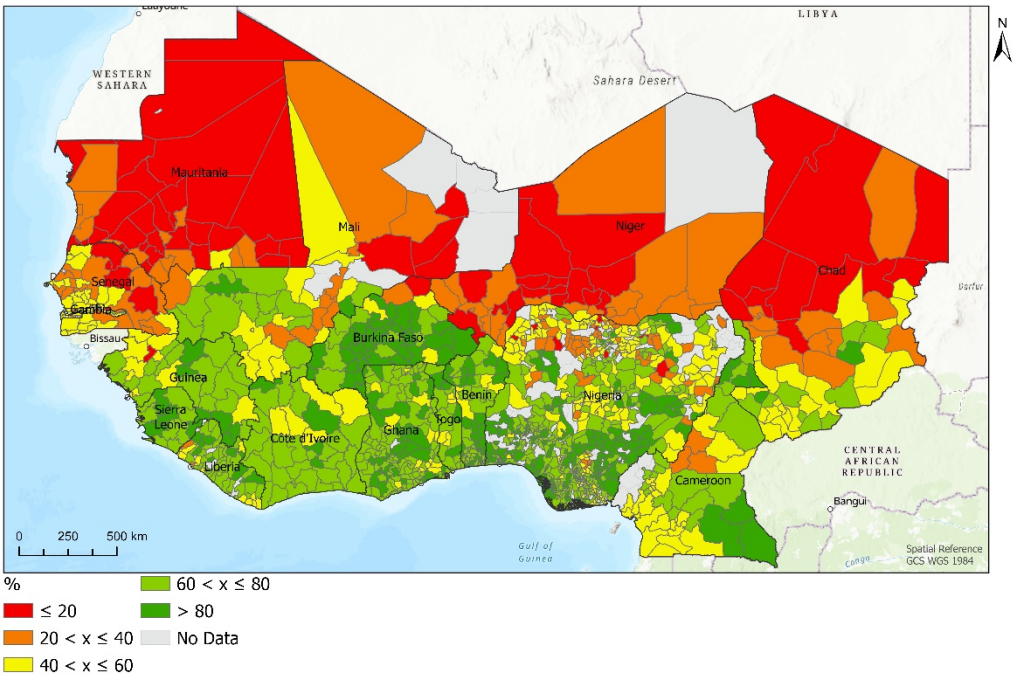

Figure II.21. Women currently employed (%).

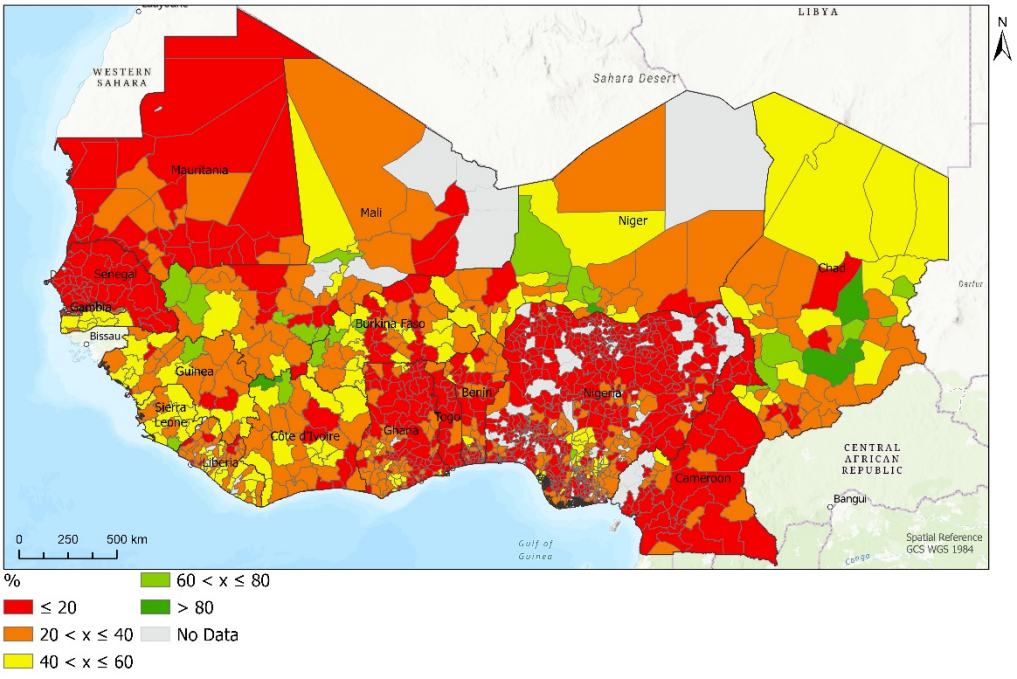

Figure II.22. Women who own a house, either alone or jointly (%).

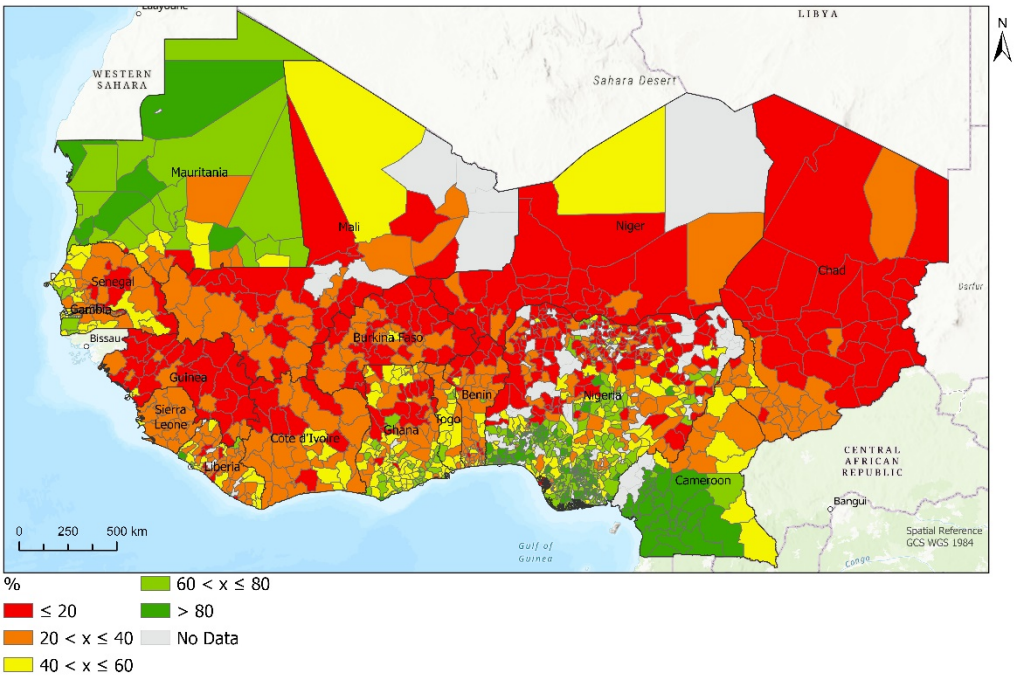

Figure II.23. Women literacy (higher than secondary or can read part or whole sentences) (%).

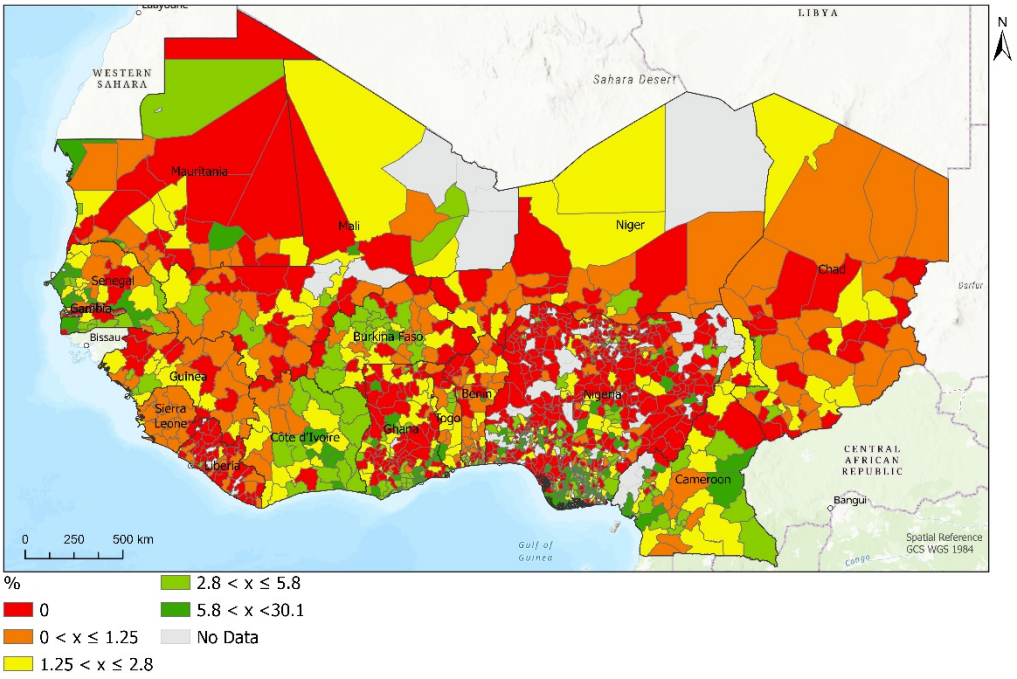

Figure II.24. Women who access all the three media (TV, radio, newspaper) on a weekly basis (%).

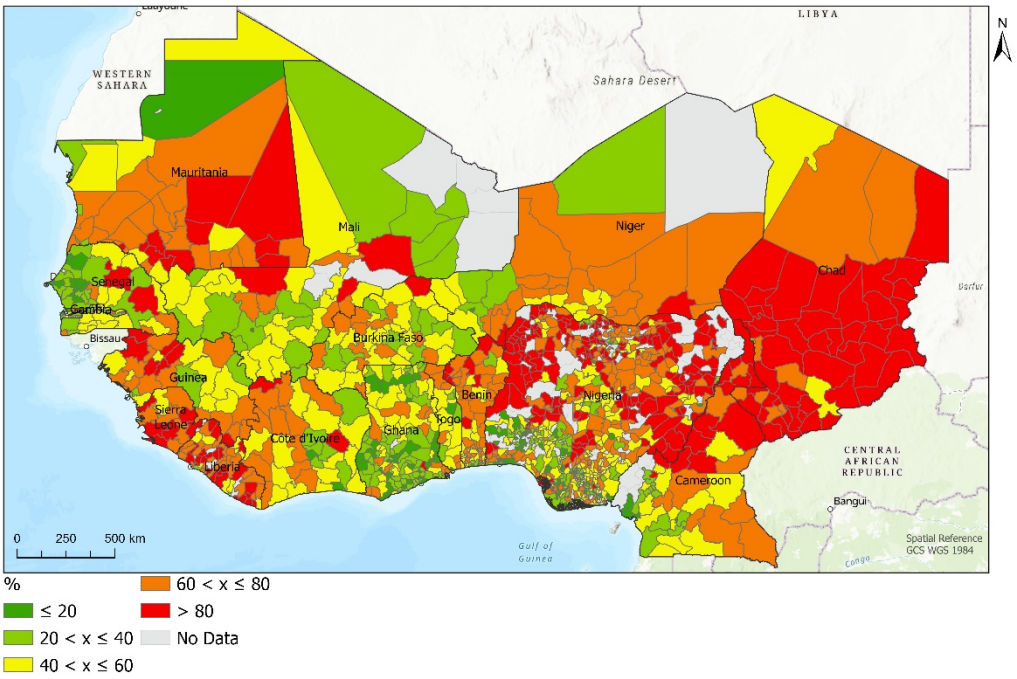

Figure II.25. Women who access none of the three media (tv, radio, newspaper) at least once a week (%).

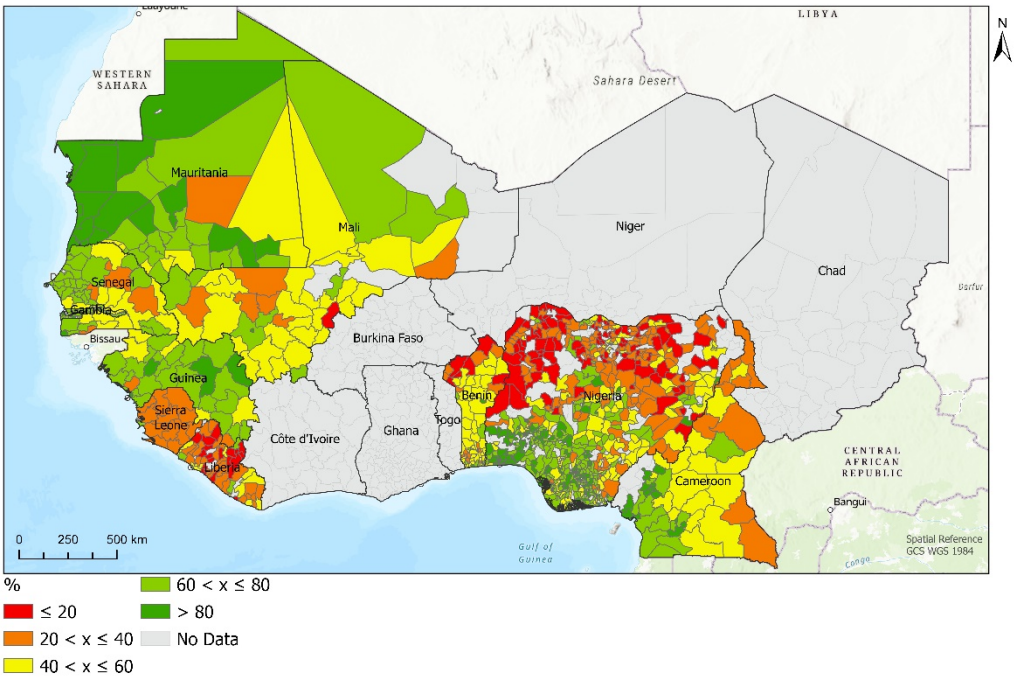

Figure II.26. Women who own a mobile phone, either alone or jointly (%).

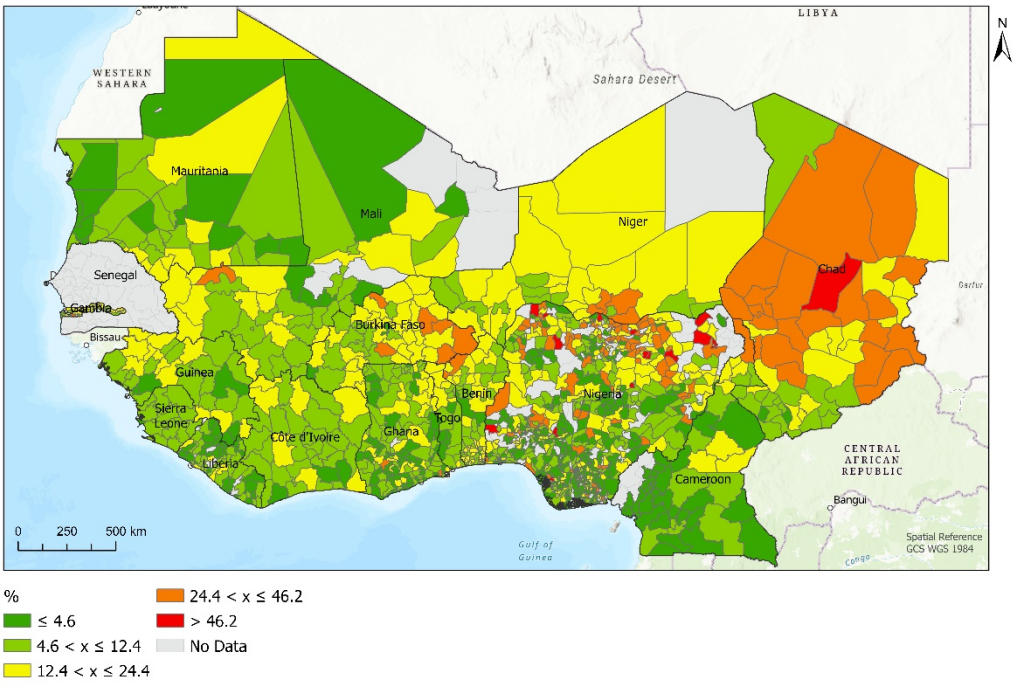

Figure II.27. Women with BMI < 18.5 kg/m2 (underweight) (%).

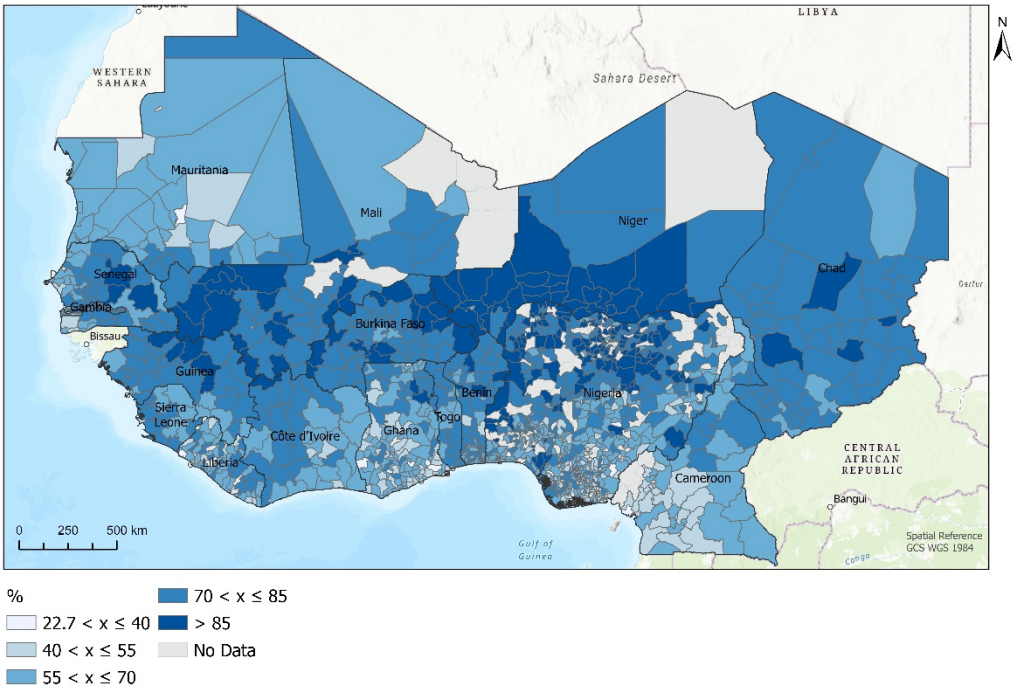

Figure II.28. Women who are married or in a union (%).

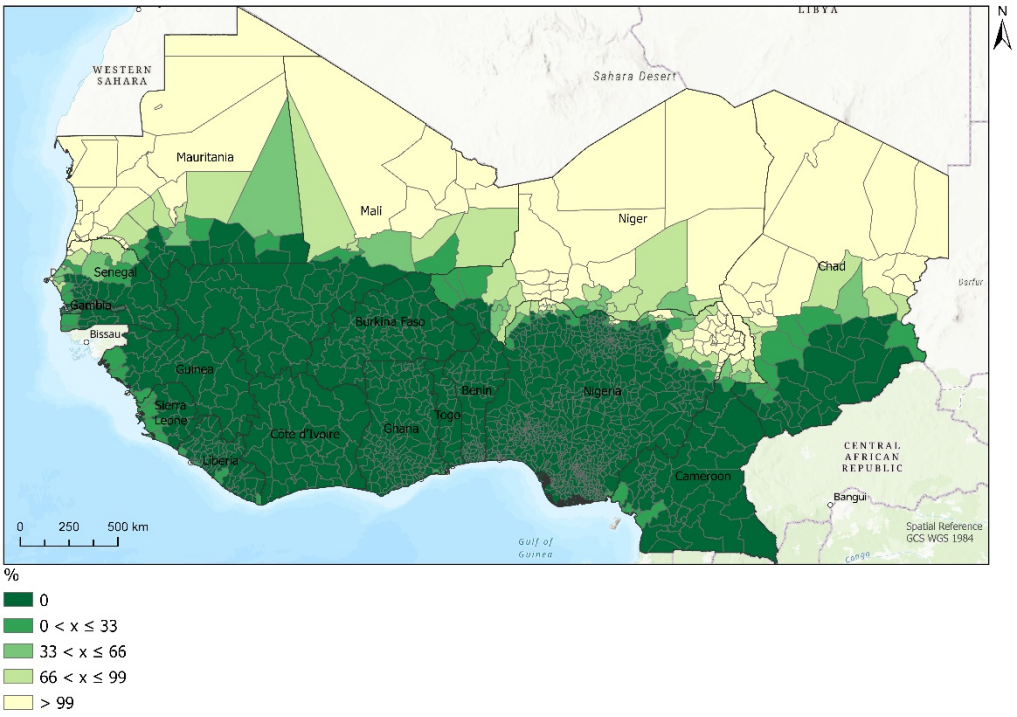

Figure II.29. Arid area (%).

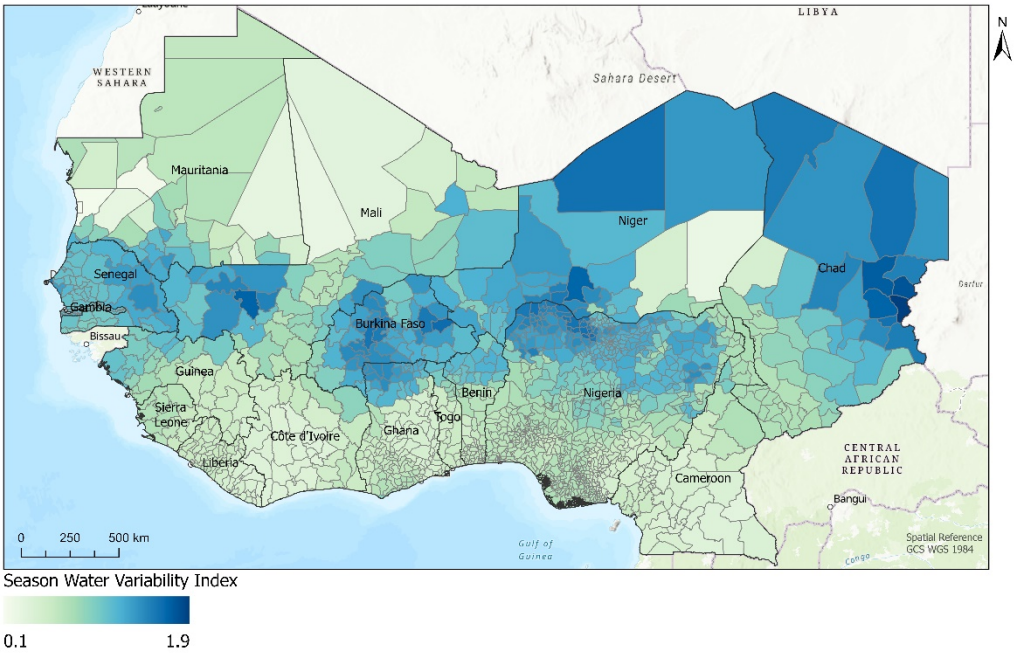

Figure II.30. Seasonal water supply variability index.

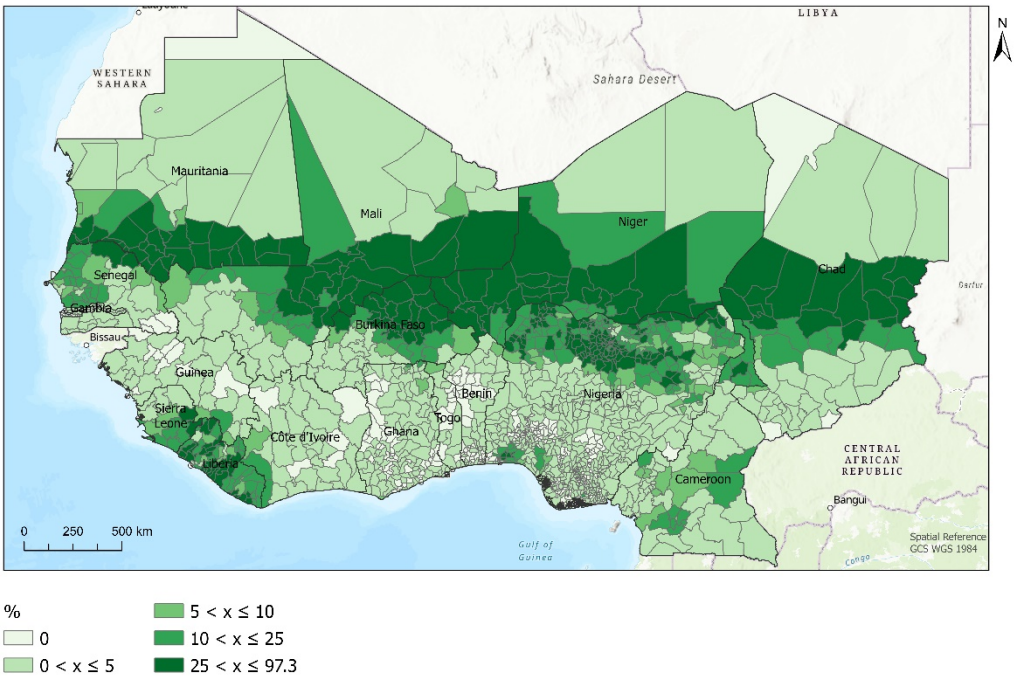

Figure II.31. Grassland coverage (%).

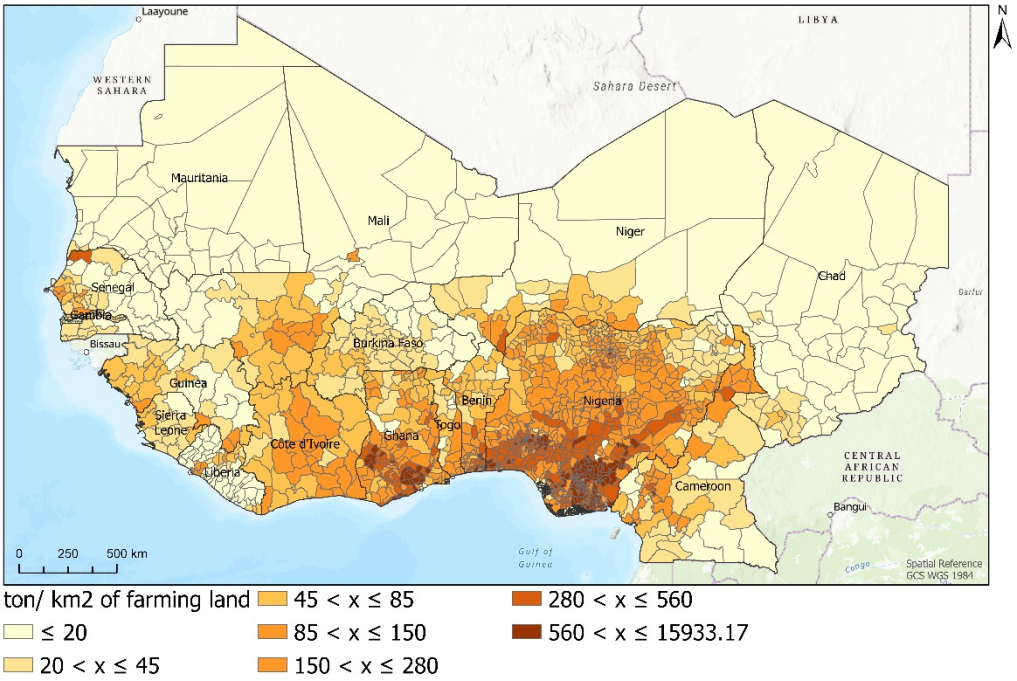

Figure II.32. Crop production per km^2 (ton/km^2).

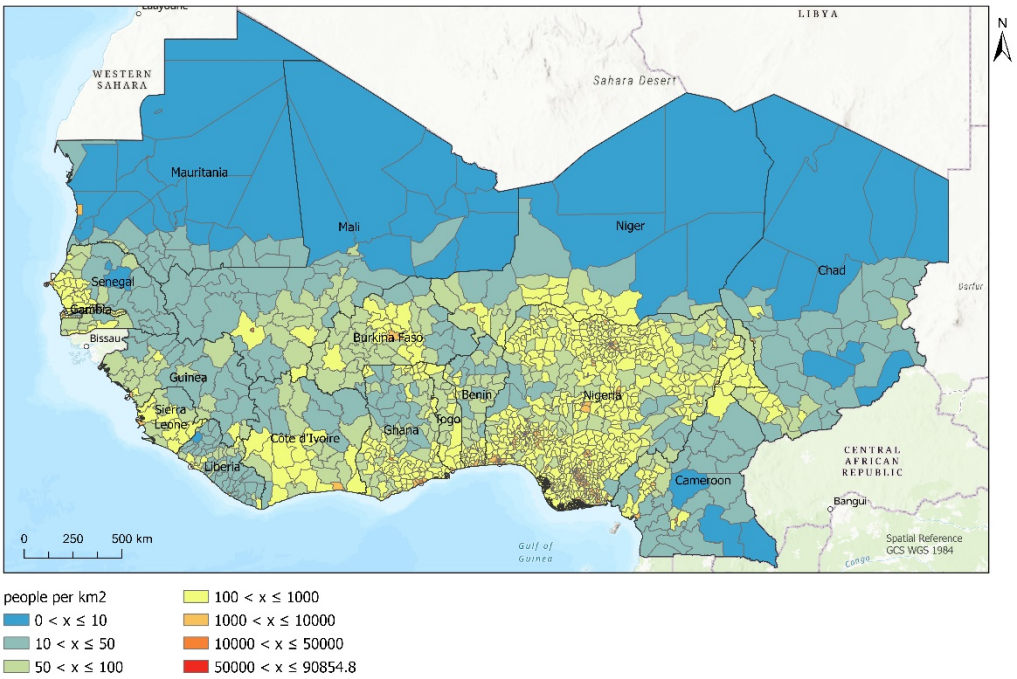

Figure II.33. Population density (people per km^2 of land area).

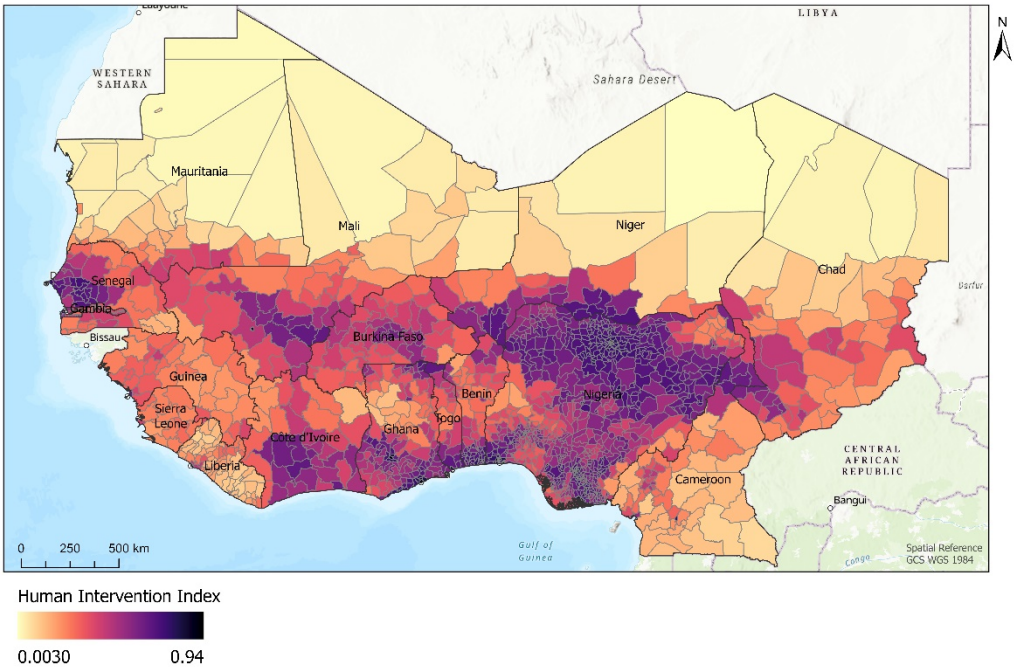

Figure II.34. Human modification index.

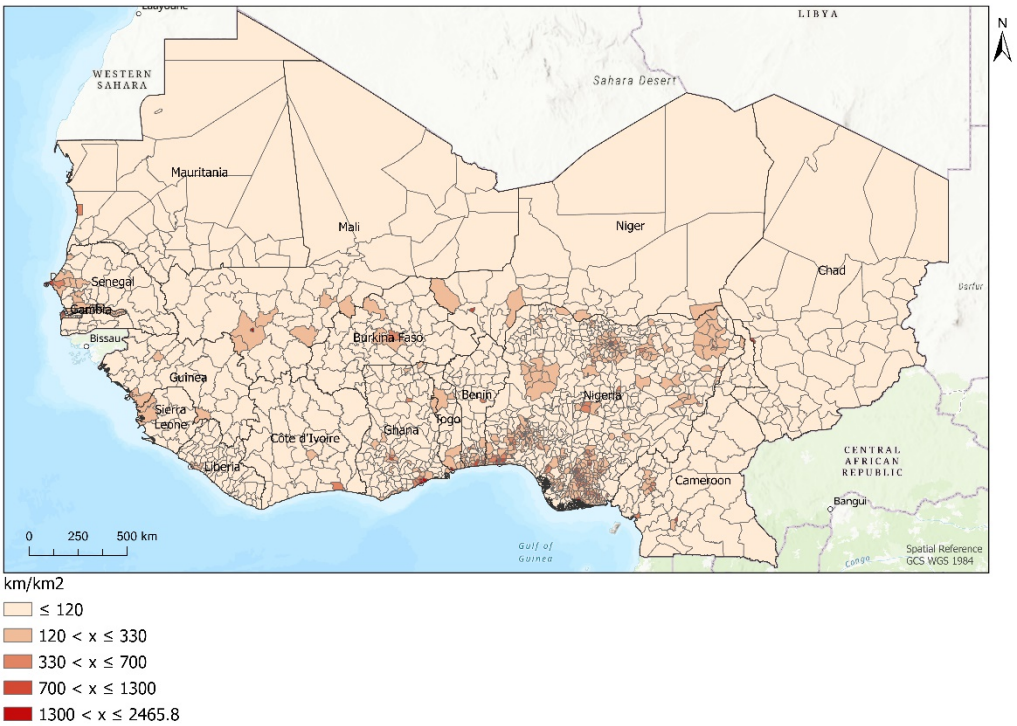

Figure II.35. Road density (roads per km^2 of land area).