Scientific African 29 (2025) e02813

Contents lists available at ScienceDirect

Scientific African

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

Bayesian geo-additive modeling of zonal level crop production in Ethiopia

Yidnekachew Mare [a,c,*], Temesgen Zewotir [b], Denekew Bitew Belay [a,d]

[a] *Department of Statistics, College of Science, Bahir Dar University, Bahir Dar P.O. Box 79, Ethiopia*
[b] *School of Mathematics, Statistics and Computer Science, College of Agriculture Engineering and Science, University of KwaZulu-Natal, Durban, South Africa*
[c] *Department of Statistics, College of Natural Sciences, Arba Minch University, Arba Minch, Ethiopia*
[d] *Department of Statistics, University of Pretoria, Pretoria, South Africa*

ARTICLE INFO

Editor: DR B Gyampoh

Keywords:
Bayesian inference
Geo-additive
Mixed models
Spatial dependence
Crop production
Zonal administrative level

ABSTRACT

Crop production plays an important role in global food security, economic stability, and sustainable development, so it is important to identify covariates that linearly and nonlinearly affect it to ensure sustainable food security and economic stability. In this study, we have used a Bayesian geo-additive mixed model to analyze the spatially structured agricultural sample survey data of eight years (Meher seasons from 2012/13 to 2019/20) collected annually by the Central Statistics Agency of Ethiopia (the current Ethiopian Statistical Service). The posterior estimates of the linear fixed effects showed that the proportion of farmers preventing soil erosion, the proportion of educated farmers, the percentage of crop damage, and the number of oxen all have a significant negative effect, while the proportion of farmers who practice pure agriculture and the area used have a significant positive effect on log crop production per household in the zone. The posterior estimates of the non-linear fixed effects showed that year, the proportion of female farmers, the proportion of farmers who practice other agriculture, the proportion of farmers who used broadcast sowing, household age, farmer association crop production, and UREA fertilizer used have significant non-linear effects on log crop production. Pure agricultural farming, cluster farming, farmers' associations, and UREA fertilizer usage are recommended to increase crop production at the zone level. To attain the main objective of this study, we considered only the spatial structure or dependency of the sample survey data.

Introduction

Crop production is important for global food security, serving as a primary source of food and essential nutrients. It also contributes to economic stability by providing income for smallholder farmers and employment for rural youth [1,2]. Furthermore, crop production supports sustainable development through efficient management of natural resources and enhances the resilience of agricultural systems to climate change [3,4]. Additionally, it plays a significant role in trade and market access, which affect exports and market dynamics [5].

Also in Sub-Saharan countries, crop production is the primary source of food security and income for smallholder private farmers [6–8]. In Ethiopia, smallholder private farmers significantly contribute to the majority of Meher (which is the main cropping season in

---

* Corresponding author at: Department of Statistics, College of Science, Bahir Dar University, Bahir Dar P.O. Box 79, Ethiopia.
E-mail address: yidnekachew.mare08@gmail.com (Y. Mare).

https://doi.org/10.1016/j.sciaf.2025.e02813
Received 14 April 2025; Received in revised form 23 May 2025; Accepted 10 June 2025

Ethiopia) season's crop production, playing an important role in household and national food security and the economy [9–11]. To ensure food security and economic sustainability, it is essential to have well-organized and managed crop production at the farmer, zone, and county levels. This approach will help enhance the productivity and efficiency of crop production by maximizing the utilization of limited resources and inputs [12–20].

Identifying covariates significantly affecting crop production and implementing the findings are important for a country to have a sustainable source of food security and economy [21–26]. For this, we can use linear or non-linear (mixed) statistical models [27–31]. The linearity assumption between the response variable and covariates in a linear mixed model may not hold for all covariates, and to relax the linearity assumption, we can use an additive mixed model [30,32–35].

If the spatial information of the area units (zones) from which the dataset was collected is available, modeling the variability in the response variable as a function of covariates and random effects that model residual variation unexplained by covariates but contributed by spatial variations will improve the predictive performance of the model. So, we extend the additive mixed model by incorporating the spatial information into the model analysis [36].

According to Waldo Tobler's first law of geography, observations from nearby area units are more similar or have a relatively stronger spatial autocorrelation than observations from distant area units [37]. In addition to repeated observations of the same area unit (zone), the spatial structure of the dataset leads to spatial dependence or autocorrelation of observations, which needs spatial analysis that addresses the dependence of observations (spatial structure); otherwise, it leads to spurious results [38,39].

So, to identify the spatial pattern of crop production across zones of the country, Ethiopia, and estimate the linear and non-linear effects of the covariates, we use the Bayesian modeling approach, Latent Gaussian Models (LGMs), which is easily analyzed using the Integrated Nested Laplace Approximation (INLA) method [39–42].

The main objective of this study is to identify the set of covariates that have significant linear and nonlinear effects on crop production and identify the spatial pattern of crop production using a Bayesian geoadditive mixed model that considers the spatial autocorrelation between observations from nearby zones. For this purpose, different packages (like "rgdal," "spdep," and "R-INLA") in R (4.3.2, updated version) software are used.

Study area and data

The study areas used in this study are all ninety administrative zones from all ten demonstrative regions in Ethiopia, where agricultural sample surveys are annually conducted by the Central Statistics Agency (CSA) of Ethiopia.

Eight-year (from 2012/13 to 2019/20) Meher seasons agricultural survey data from smallholder private farmers is aggregated at zonal level using weighted average (mean), and we used a natural log transformation of the response variable, the mean crop production per zone per household, following the results of a Box-Cox transformation.

Methodology

Spatial statistics is a very broad collection of methods and techniques of visualization, exploration, and analysis applied to spatial data with the interest of studying the spatial pattern of the interest variable over spatial area units. Spatial data are stochastic process realizations of attributes indexed by area unit (zone), and if there are m non-overlapping spatial area units (zones), the observed attribute in area unit j is given by [38,39]:

$Z(s_j) : s \in D \subset R^k \quad j = 1, 2, 3, \ldots,$ m (m = 90 zones in our case)

where $Z(s)$: is a characteristic of the interest variable observed at spatial location s_j

s: is the spatial location from where data is observed (each zone)

D: is the spatial domain or area of interest (the study area, Ethiopia)

R^k: is k-dimensional Euclidean space (2-dimensional in our case).

Depending on the type of problem and characteristics of the domain D, there are three types of spatial data: areal (lattice) data, geostatistical data, and point pattern. For this study, we use areal (lattice) data from irregularly shaped administrative zones that share non-zero-length common boundaries [38,39].

Following Waldo Tobler's first law of geography, we expect that observations from nearby zones have a relatively stronger spatial autocorrelation than observations from distant zones, and this leads to spatial dependency or correlation that decreases with the increase in separation distance (h) between zones [37,38,41].

A spatial weights matrix is a mathematical construct used in spatial analysis to represent the spatial structure of a dataset (i.e., the spatial dependency between different zones) and to compute variogram and spatial autocorrelation measures like Moran's I and Geary's C. The spatial weights matrix is an m × m matrix, where m is the number of zones, and each element w_{ij} represents the spatial relationship between zones s_i and s_j.

The weights w_{ij} can take various forms, and we considered four different schemes: radial distance weights (186 km threshold), k-nearest neighbor weights ($k = 5$), and boundary-based or contiguity weights (rook contiguity weights and queen contiguity weights) for this study [38,43–47].

To measure and test the spatial stationarity of the process and spatial autocorrelation of observations (of interest variable) from nearby zones, we can use a variogram (semi-variogram), global and local Moran's I and Geary's C, Getis and Ord local statistics, and a Moran scatterplot [38,45,46,48]:

Bayesian inference is a flexible and powerful statistical method that uses Bayes' theorem to update the probability of a hypothesis based on new evidence, allowing for the incorporation of prior knowledge and providing a comprehensive probabilistic framework for data interpretation. Bayes' theorem relates the conditional and marginal probabilities of random variables, and the target is to estimate the joint posterior distributions [49–52].

Integrated Nested Laplace Approximation (INLA) is a computational method for Bayesian inference in latent Gaussian models (like spatial models), offering a fast and accurate alternative to traditional Markov Chain Monte Carlo (MCMC) methods. Unlike MCMC, which focuses on estimating the highly multivariate joint posterior distribution, INLA estimates the univariate marginal posterior distribution of latent fields and hyperparameters. INLA's computational efficiency and superiority stem from its use of sparse representations of high-dimensional precision matrices in latent Gaussian models and its combination of Laplace approximations and nested integration strategies. These advantages make INLA widely applicable in various fields, and it is implemented in the R-INLA package [39,40,53–57].

The general form of a geo-additive mixed model for a normally distributed response variable is given by (1):

$$\boldsymbol{y} = \boldsymbol{X}_i\boldsymbol{\beta} + \sum_{k=1} f_k(x^*{}_{ki}) + f_{struct}(\cdot) + f_{unstruct}(\cdot) \tag{1}$$

where $\boldsymbol{y}$ is a vector of the response variable (n × 1)

$\boldsymbol{X}_i$ is the i^{th} row of the known design matrix ($\boldsymbol{X}_{n\times c}$) of fixed effects
$\boldsymbol{\beta}$ is a vector of unknown linear fixed effects (c × 1)
$f_k(\cdot)$ is unknown non-linear smooth function of the covariate x^*_k
$f_{struct}(\cdot)$ is a structured (spatially correlated) random effect (u_j)
$f_{unstruct}(\cdot)$ is an unstructured (spatially uncorrelated) random effect (v_j).

The structured and unstructured spatial random effects are modeled using the 'BYM' model by Besag, York, and Mollié, and for the i^{th} observation of the response variable, the model is re-expressed by (2) as follows [39,53,58]:

$y_i = \boldsymbol{X}_i\boldsymbol{\beta} + \sum_{k=1} f_k(x^*{}_{ki}) + u_j + v_j + \varepsilon_i$(2) where $u_j|\boldsymbol{u}_{-j}, \ \tau_u \sim N\left(\bar{u}_{n_j}, \ \frac{1}{\tau_u n_j}\right)$ … iCAR model in which the value of the spatial effect u_j at zone s_j is conditionally dependent on the neighboring spatial effect values $\boldsymbol{u}_{-j}$

$\bar{u}_{n_j} = \frac{1}{n_j}\sum u_{-j}$ where n_j is the number of neighbors of zone s_j

$v_j|\tau_v \sim N\left(0, \frac{1}{\tau_v}\right)$ … i.i.d. Gaussian model

$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ … random error identically and independently normally distributed.

The terms τ_u and τ_v are the precisions (variance inverse) of the structured and unstructured spatial random effects, respectively, and σ_ε^2 is the variance of the random error term.

For this study, the Gaussian priors (mean of zero and a precision of 0.001) are used for fixed effect and random walk parameters. For the spatial random effects, we used log gamma with parameters of (1, 0.001) for the spatially structured random effect and (1, 0.01) for the spatially unstructured random effect [39,54,58].

The basic assumptions for geo-additive mixed models are extensions of the assumptions for linear (additive) mixed models, and the spatial process is often assumed to be stationary (i.e., the statistical properties do not change over space) [31,32,38,59–62].

For model assessment (how they are fitting the data) and comparison (among the competing models), we can use different criteria

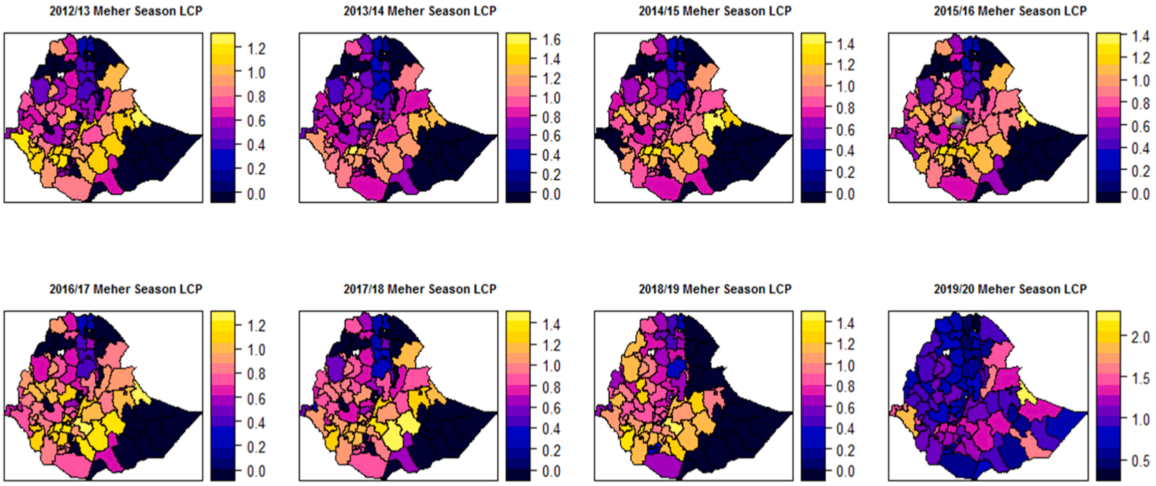

Fig. 1. Observed Meher season crop production by zones from 2012/13 to 2019/2020.

values computed using R-INLA, like conditional predictive ordinate (CPO), predictive integral transform (PIT), and Kullback-Leibler divergence (kld) for model assessment and marginal likelihood (ML), deviance information criterion (DIC), and widely applicable Bayesian information criterion (WAIC) for model selection [41,58,63–72].

Results and discussion

The overall mean and standard deviation of crop production across all zones for the study period are 0.850 and 0.267 quintals, respectively. The minimum and maximum crop production observed are 0.208 and 2.170 quintals, respectively. The first, second, and third quartiles of crop production are 0.675, 0.845, and 1.036 quintals, respectively, and we can see half of the farmers produced less than the overall mean crop production.

The linearity assumption of the functional relationships between the response variable and the covariates is assessed, and the result showed that nine of the covariates have a significant nonlinear relationship (large effective degrees of freedom (edf) values with small p-values (< 0.05)) with crop production.

Fig. 1 shows the spatial distribution of crop production for each zone for each Meher season, and from the figure we can see that there is clustering (similarity of crop production from nearby zones than from distant zones) of crop production, suggesting possible spatial dependency of crop production.

Fig. 2 shows the stationarity of the random process crop production (a red line fitted to the scatter plot and cloud of squared difference, top left and right panels, respectively, and the bounded fitted theoretical variograms, bottom panel). From the empirical and theoretical variograms, bottom panel, we can also see the degree of spatial dependence between pairs of crop production for different separation distances, and those observations from neighboring zones have similarity (less variance) and a high degree of spatial autocorrelation, while those from long-distance zones have high dissimilarity (high variance) and a low degree of spatial autocorrelation [73–75].

We used Moran's I and Geary's C statistics and Moran's scatterplot to test and confirm the spatial dependency of crop production,

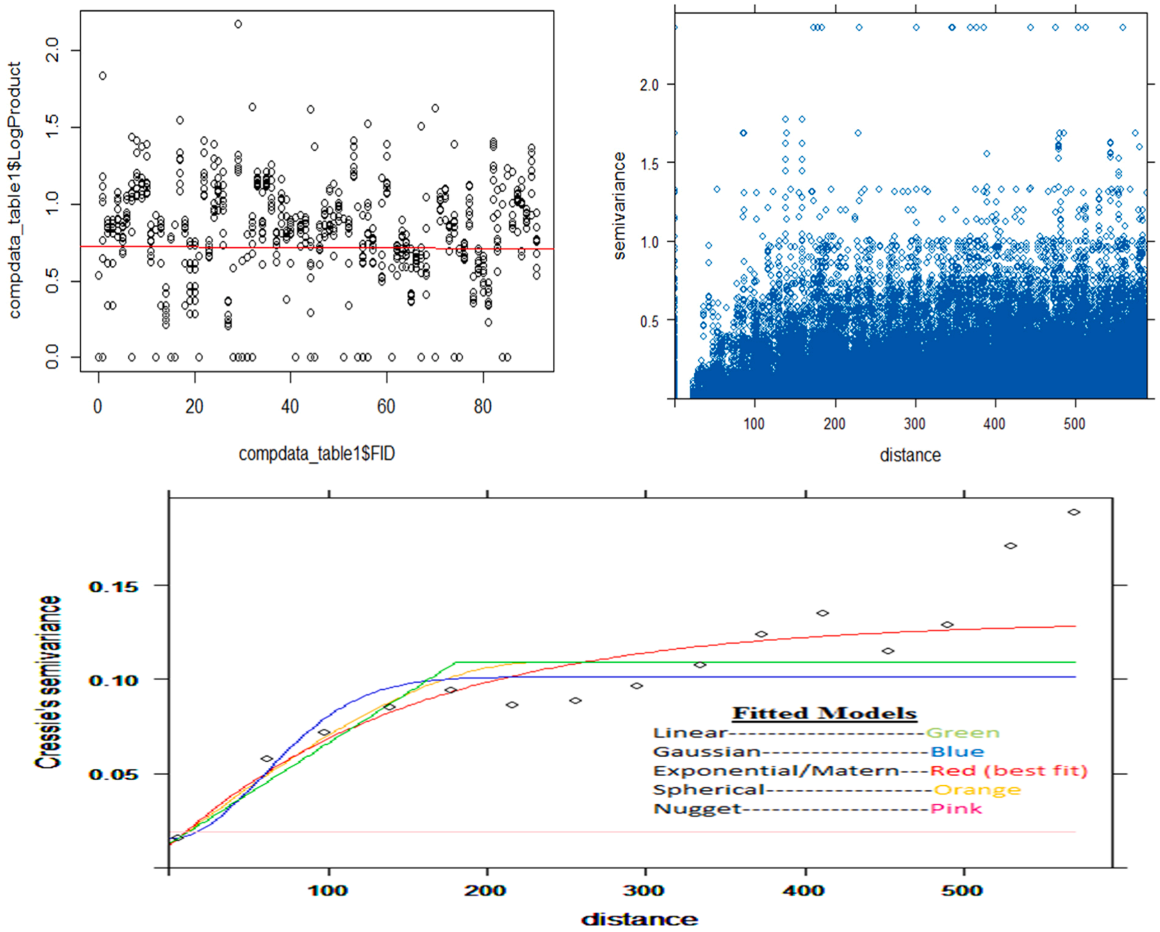

Fig. 2. Scatter plot and cloud variogram (top left and right panels, respectively) and empirical and theoretical variograms (bottom panel) of crop production.

suggested and shown in Figs. 1 and 2. We considered four different types of neighborhood schemes: rook-based contiguity, queen-based contiguity, k-nearest-based neighbors, and distance-based neighbors.

The spatial dependency (autocorrelation) of the crop production from nearby zones for each Meher season is computed using Moran's I and Geary's C under the randomization and normality assumptions, and the results are summarized and presented under Table 1. The estimated values of Moran's I and Geary's C are statistically significant, indicating there is a positive spatial autocorrelation of crop production from nearby zones.

Fig. 3 shows the Moran scatterplot of the 2019/20 Meher season's crop production using the Rook/Queen scheme (Moran's I = 0.528, p-value < 0.000), and we can see that the dominant points are found in the upper right and lower left quadrants compared to the other two quadrants, indicating there are positive spatial associations such that higher (lower) crop productions are surrounded by higher (lower) crop productions.

From the global spatial autocorrelation statistics (Moran's I and Geary's C) and the Moran scatterplot, we have seen that there are spatial dependencies (clusters) of crop production from nearby zones, but they didn't tell us where these clusters are.

Local Moran's I is computed for each zone and mapped to identify local similarity or cluster of crop production. Fig. 4 shows zones with similar crop production (zones with light-green, green, and dark-green colors) based on the four neighborhood schemes for the 2019/20 Meher season.

To verify the importance of relaxing the linearity assumptions and incorporating random effects, three different models (linear fixed model (lm), additive fixed model (af), and additive mixed model (am)) are fitted to the same dataset. Fig. 5 shows the maps of the fitted values (top panel) and their corresponding residual values (bottom panel) of the fitted models, and we can see that the additive mixed model is best performing in predicting the observed crop production, justifying the importance of considering non-linear relationships and zones as random effects in the model.

Bayesian additive models without random effects, with random effects (assuming independent zones), and Besag, York, and Mollie for queen/rook and distance-based neighborhood schemes were fitted, and all the information criteria values under Table 2, except the marginal log-likelihood, suggested the Besag, York, and Mollie model for queen/rook neighborhood scheme best fits the dataset [76].

The Besag-York-Mollie model is fitted to the data, and the posterior point and interval estimates of model parameters are shown in Table 3.

Zones with a higher proportion of educated farmers, a higher proportion of farmers who prevented erosion, a higher percentage of crop damage, and a higher number of oxen have less crop production, while zones with a higher proportion of farmers practicing pure agriculture and with more cropland area used have more crop production. The point posterior mean estimates of linear fixed effects show that covariates such as the average area used per hectare per household and the proportion of farmers who practicing pure agriculture have a positive and significant effect on crop production. This indicates that cropland and focusing on crop farming are the most important input factors. So, to increase crop production in zones, farmers should be encouraged to use cluster single crop farming to maximize effective usage of their cropping land [61,62,77–82].

The decreasing effects of education and the number of oxen may be due to the fact that farmers with higher educational levels are more likely to engage in various personal (such as other business ventures), social, and governmental activities. Additionally, households with more oxen are more likely to benefit from the growing demand for animal products as a source of food and income, as opposed to traditional labor- and capital-intensive crop farming [83–86].

Covariates, the proportion of farmers who practice pure agriculture and the average area used per hectare per household, were also found to have positive and significant effects on crop production under the linear mixed model and additive mixed model analyses of

Table 1
Moran's I and Geary's C based on four neighborhood schemes.

Meher Season	Statistic	Spatial Weights								
		Rook/Queen Contiguity			K (5)-Nearest			Distance (Within 186 km)		
		Sample Estimate			Sample Estimate			Sample Estimate		
		Estimate	Expectation	Variance	Estimate	Expectation	Variance	Estimate	Expectation	Variance
2012/13	Moran's I	0.340	−0.011	0.004	0.330	−0.011	0.004	0.275	−0.011	0.003
	Geary's C	0.670	1	0.005	0.629	1	0.004	0.707	1	0.003
2013/14	Moran's I	0.354	−0.011	0.004	0.350	−0.011	0.004	0.305	−0.011	0.003
	Geary's C	0.651	1	0.005	0.618	1	0.004	0.682	1	0.003
2014/15	Moran's I	0.340	−0.011	0.004	0.312	−0.011	0.004	0.275	−0.011	0.003
	Geary's C	0.662	1	0.005	0.657	1	0.004	0.709	1	0.003
2015/16	Moran's I	0.339	−0.011	0.004	0.339	−0.011	0.004	0.306	−0.011	0.003
	Geary's C	0.644	1	0.005	0.627	1	0.004	0.682	1	0.003
2016/17	Moran's I	0.344	−0.011	0.004	0.343	−0.011	0.004	0.328	−0.011	0.003
	Geary's C	0.644	1	0.005	0.616	1	0.004	0.650	1	0.003
2017/18	Moran's I	0.358	−0.011	0.004	0.358	−0.011	0.004	0.348	−0.011	0.003
	Geary's C	0.653	1	0.005	0.603	1	0.004	0.622	1	0.003
2018/19	Moran's I	0.528	−0.011	0.004	0.448	−0.011	0.004	0.459	−0.011	0.003
	Geary's C	0.467	1	0.005	0.486	1	0.004	0.515	1	0.003
2019/20	Moran's I	0.528	−0.011	0.004	0.448	−0.011	0.004	0.459	−0.011	0.003
	Geary's C	0.467	1	0.005	0.486	1	0.004	0.515	1	0.003

p-value = 0.000 for all Meher seasons, neighborhood schemes, and assumptions

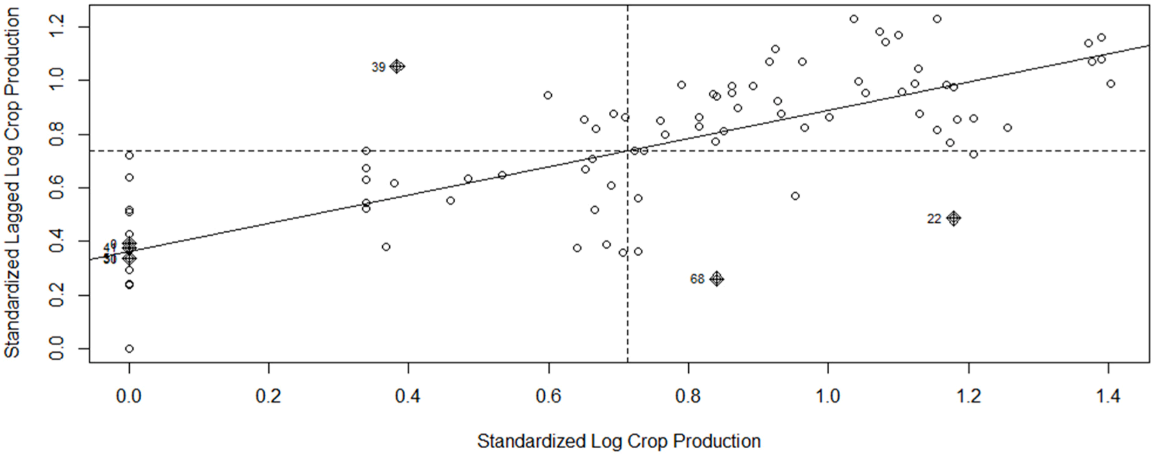

Fig. 3. Moran scatterplot of the 2019/20 Meher season's crop production.

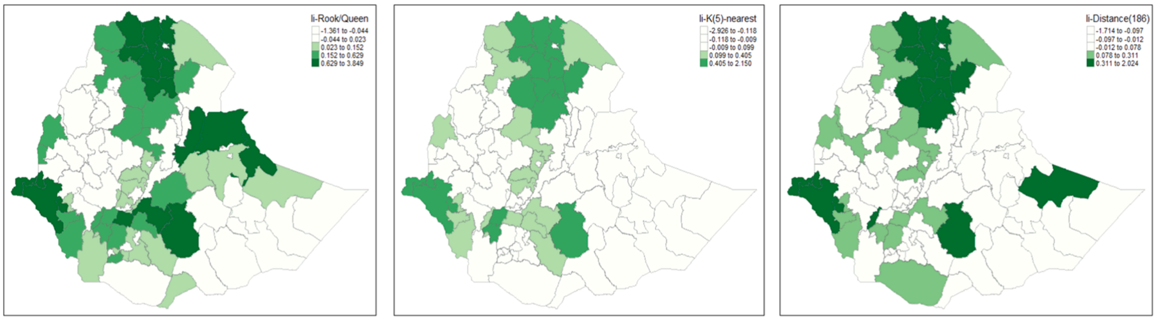

Fig. 4. Local Moran's I plot of the 2019/20 Meher season based on four neighborhood schemes.

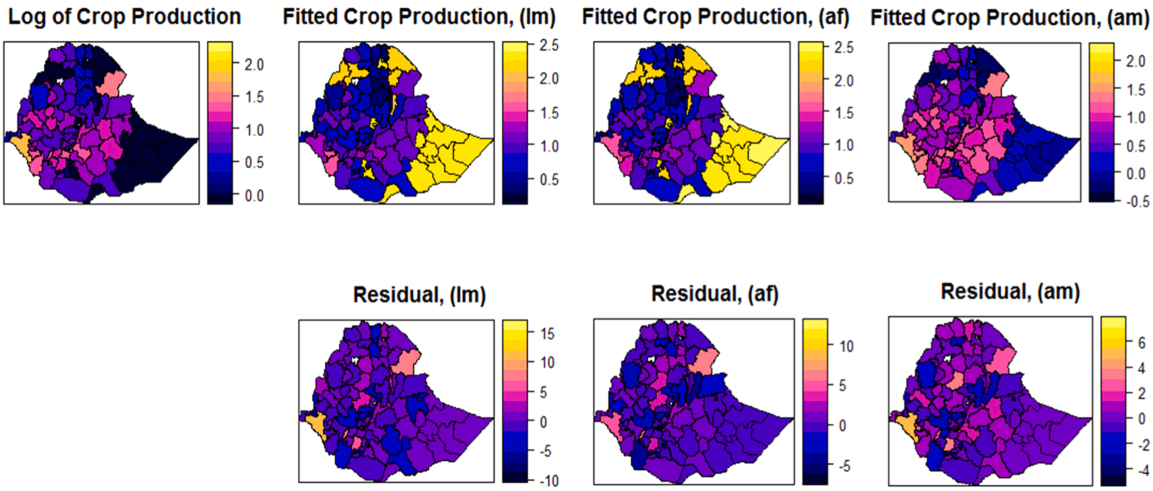

Fig. 5. Maps of observed and fitted (top panel) and corresponding residual (bottom panel) values from linear (lm), additive fixed (af), and additive mixed (am) models.

the previous studies. The covariate proportion of educated farmers was also found to have a negative and significant effect on crop production under additive mixed model analysis but was insignificant under linear mixed model analysis of the previous studies [61, 62].

The covariates proportion of farmers who get credit service and proportion of farmers who hold only crop are insignificant as they were under linear mixed model analysis, but they were found to have negative and significant effects on crop production under

Table 2
Summary of information criteria values for the Bayesian models.

Fitted Model (Zone as random effects)	CPO	DIC	WAIC	ML (Gaussian)
Without random effects	−521.4705	−1102.35	−1064.08	460.32
With random effects, assuming iid	−637.0227	−1394.53	−1340.49	528.90
With structured and unstructured random effects, for Queen/Rook contiguity	−638.4874	−1399.49	−1345.15	470.96
With structured and unstructured random effect, for Distance contiguity	−637.2573	−1395.73	−1341.57	444.83

Table 3
Posterior means, standard deviations, and 95 % credible intervals of fixed effects and model hyperparameters.

Fixed effects:							
	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	0.701	0.228	0.249	0.703	1.146	0.703	0
Educ_Prop	−0.002	0.001	−0.004	−0.002	−0.001	−0.002	0
PureAgr_Prop	0.004	0.001	0.003	0.004	0.005	0.004	0
CropHold_Prop	−0.001	0	−0.002	−0.001	0	−0.001	0
BothHold_Prop	−0.001	0.001	−0.003	−0.001	0.001	−0.001	0
PrivateOwner_Prop	0.002	0.001	−0.001	0.002	0.004	0.002	0
Credit_Prop	−0.001	0.001	−0.002	−0.001	0	−0.001	0
Advice_Prop	0	0.001	−0.001	0	0.001	0	0
CropRotat_Prop	0	0.001	−0.001	0	0.001	0	0
Extension_Prop	0.001	0.001	−0.001	0.001	0.002	0.001	0
Irrigation_Prop	0	0.001	−0.002	0	0.003	0	0
PrevErosion_Prop	−0.002	0.001	−0.003	−0.002	−0.001	−0.002	0
ImpSeed_Prop	−0.001	0.001	−0.003	−0.001	0.002	−0.001	0
CropDamage_Prop	0.001	0.001	0	0.001	0.002	0.001	0
MeasureDamage_Prop	0	0.001	−0.002	0	0.001	0	0
HH_Size	0.011	0.01	−0.01	0.011	0.032	0.011	0
Area	1.345	0.207	0.939	1.345	1.75	1.345	0
DA_Product	−0.001	0	−0.001	−0.001	0	−0.001	0
Imp_Seed	0	0	0	0	0	0	0
DAP_Fert	−0.001	0.001	−0.002	−0.001	0.001	−0.001	0
Other_Fert	0.001	0.001	0	0.001	0.003	0.001	0
Dam_Perc	−0.003	0.001	−0.004	−0.003	−0.001	−0.003	0
Num_OXEN	−0.151	0.037	−0.224	−0.151	−0.077	−0.151	0
FruBear_Trees	0	0	0	0	0	0	0
Random effects:							
Model hyperparameters:	mean	sd	0.025quant	0.5quant	0.975quant	mode	
Precision for the Gaussian observations	156.56	10.47	136.92	156.22	178.12	155.58	
Precision for Zone.struct	2140.39	9113.07	11.13	454.84	14,579.67	12.62	
Precision for Zone	126.53	33.05	74.4	122.23	203.61	113.9	
Adjusted R-Squared (R^2-adj)	0.923						

additive mixed model analysis of the previous studies. The covariate proportion of farmers who prevented erosion was insignificant under the linear mixed model and additive mixed model analyses of the previous studies. The covariate percentage of crop damage was found significant to affect crop production negatively under linear mixed model analysis but insignificant under additive mixed model analysis of the previous studies. The covariate number of oxen was found significant to affect crop production negatively under additive mixed model analysis but insignificant under linear mixed model analysis of the previous studies [61,62].

We also observed that the covariate irrigation is not significant in explaining crop production under this study and the previous studies using linear mixed model analysis and additive mixed model analysis, as this study mainly focuses on the Meher season's crop production, which is the main and rainy season in the country [87].

Covariates non-linearly affecting crop production are shown with their magnitude and direction in Figs. 6a and 6b, and we can see that the posterior mean effects are positive above the red dashed lines, which are at the zero posterior mean effects.

Table 4 and Fig. 7 show the summary statistics and maps of observed and fitted crop production from each zone for the study period, respectively, using the Bayesian geo-additive mixed model. From the table and maps, we can see that the fitted values are very close to the observed values, indicting the predictive power and how the model fitted the data.

Fig. 8 shows the plots of the spatial random effects considered in the model that capture the zonal-level variability in crop production not explained by the linear and non-linear effects. The spatially structured random effects indicated the spatial clustering of crop production by zones, and we can see that the western and eastern zones of the country have higher crop production compared to the northern, central, and south-central zones of the country (Fig. 8(A)). The spatially unstructured random effects (Fig. 8(B)) show the randomness or independence of crop production by zones except some zones from southern Oromia and Somali regions.

Fig. 9 shows the PIT (top panel) and CPO (bottom panel) plots, and we can see that the number of model failures to predict each observation using leave-one-out cross-validation is zero, the distribution of observations' PIT values closely follows a uniform

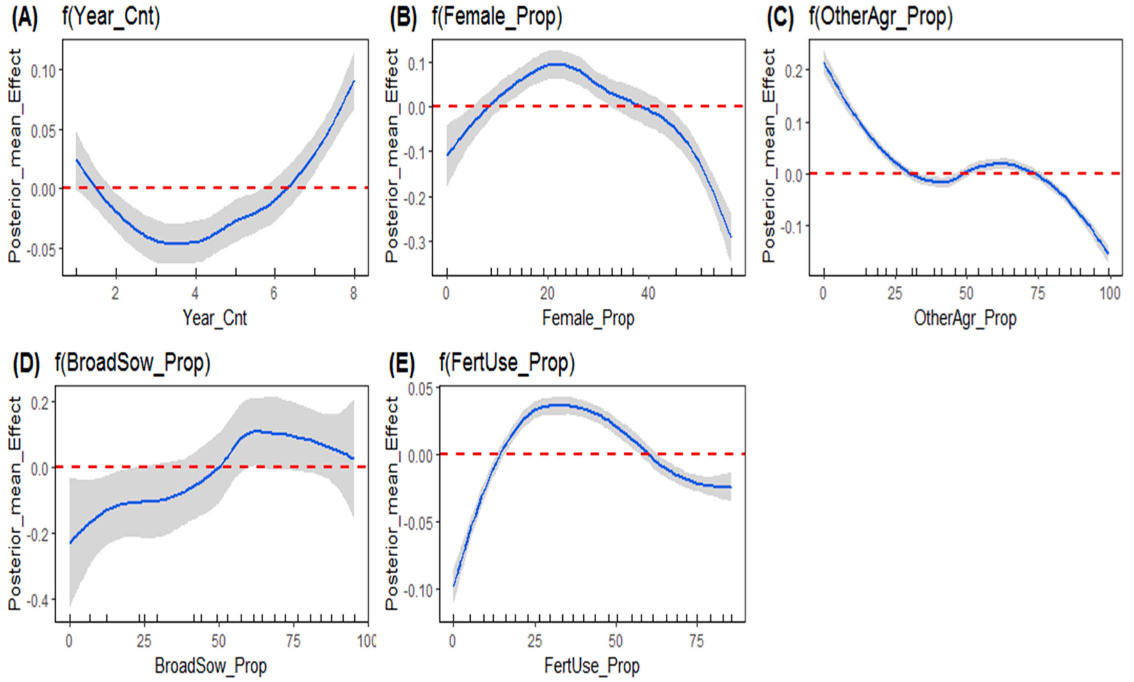

Fig. 6a. Non-linear effects; (A) Meher season, (B) proportion of female farmers, (C, D, and E) proportion of farmers who practiced other agriculture, broadcast sowing, and used any fertilizer, respectively.

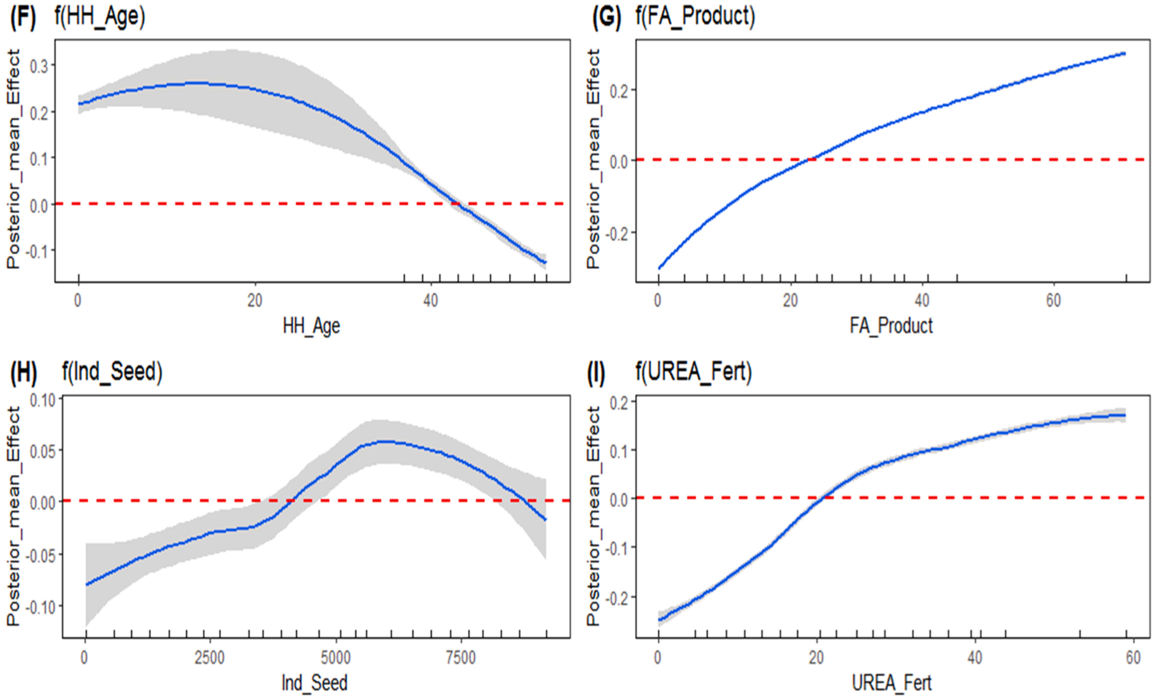

Fig. 6b. Non-linear effects; (F) household age, (G) farmers association crop production (in quintals), (H) indigenous seed used (in kg), and (I) only UREA fertilizer used (in kg).

Table 4
Summary of observed and fitted log crop production.

Log crop production	Mean	Sd	0 % (Min)	25 % (Q1)	50 % (Median)	75 % (Q3)	100 % (Max)
Observed	0.8496	0.2665	0.2083	0.6736	0.8445	1.0360	2.1698
Fitted	0.8502	0.2541	0.2237	0.6884	0.8483	1.0229	2.0372

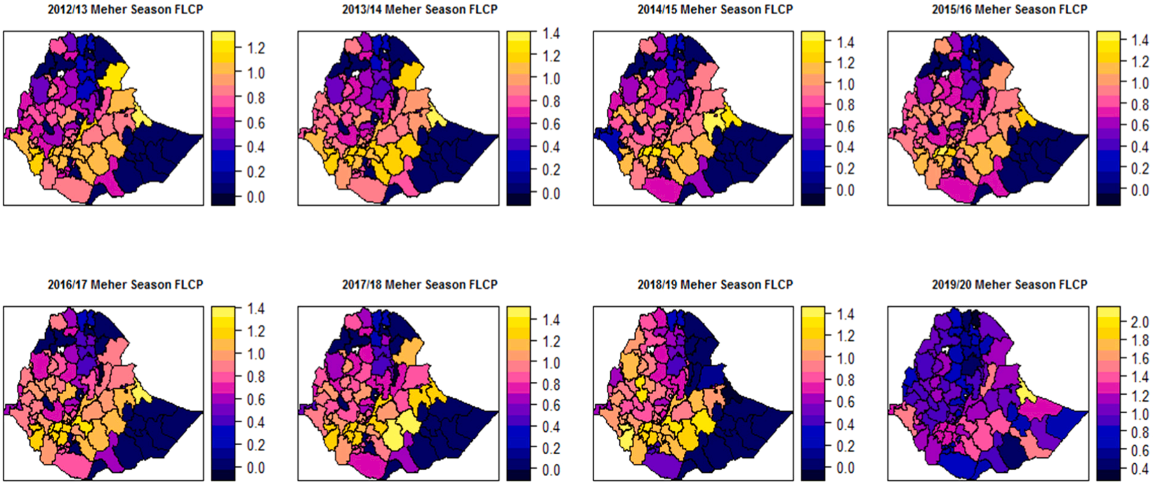

Fig. 7. Fitted Meher season crop production by zones from 2012/13 to 2019/2020 using Besag-York-Mollie model.

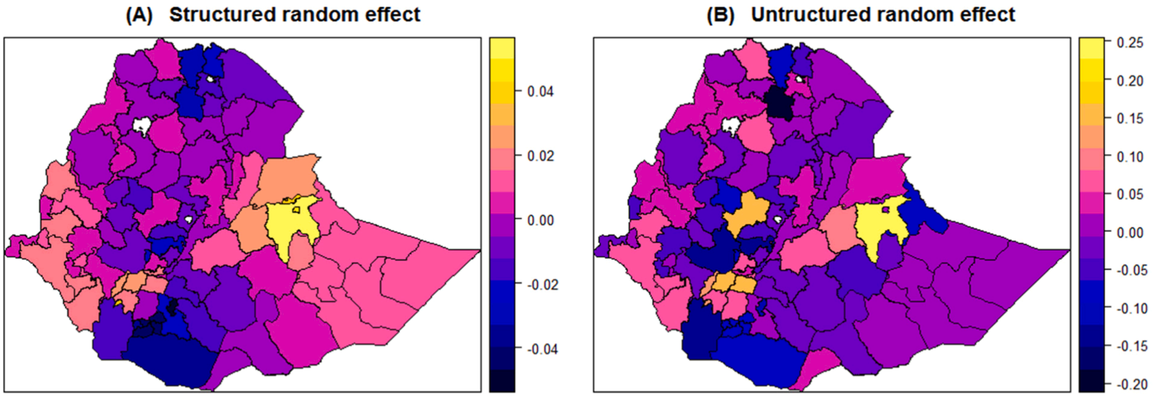

Fig. 8. Spatial random effects; (A) spatially structured, (B) spatially unstructured.

distribution between zero and one (top panel), and there are no unusual or unlikely observations to indicate outliers or extreme values (bottom panel), all of which indicated the goodness of the final model to fit the data.

Conclusion

Relaxing the linearity assumptions and incorporating spatially structured and unstructured random effects into the model improved the predictive power of the model, and the Bayesian geo-additive mixed model best fitted the data compared to the classical linear fixed, additive fixed, and additive mixed models.

Based on the posterior mean and uncertainty (95 % credible interval) estimates of linear effects, we conclude that covariates educational level, soil erosion, crop damage, and number of oxen have negative effects on crop production, while covariates pure agricultural practice and cropland (area) have positive effects on crop production. Among the linearly affecting covariates, area used and number of oxen are the most significant positively and negatively, respectively, contributing factors for crop production in the administrative zone.

The spatially structured random effects showed spatial clustering of crop production by zones, and we observed that the western

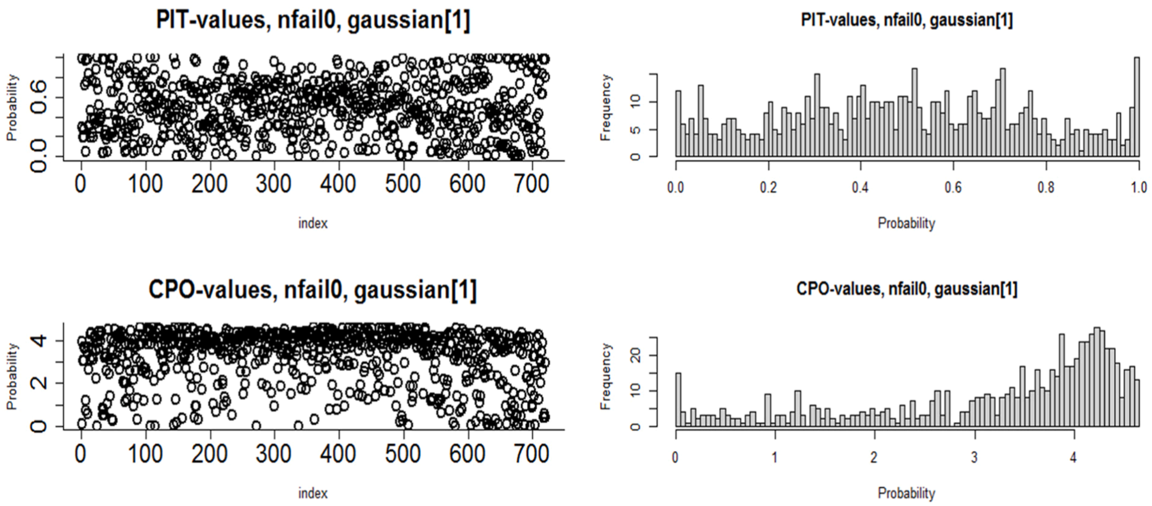

Fig. 9. PIT (top panel) and CPO (bottom panel) plots for Besag-York-Mollie model.

and eastern zones of the country are more effective in crop production than the northern, central, and south-central zones of the country.

A Bayesian geo-additive mixed model is used when the functional relationship between covariates and the response variable is not strictly linear and there are spatial dependencies between observations. If the dataset is both spatially and temporally referenced, spatial modeling has the limitation of capturing the dependency and variation of observations over time. So, the spatial and temporal nature of the dataset leads to spatial and temporal dependence of observations, which needs spatiotemporal modeling to effectively capture the complex patterns and trends by including spatial and temporal random effects and possibly their interaction effects.

CRediT authorship contribution statement

Yidnekachew Mare: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft. **Temesgen Zewotir:** Conceptualization, Investigation, Validation, Resources, Writing – review & editing, Supervision. **Denekew Bitew Belay:** Conceptualization, Investigation, Validation, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D.D. Headey, W.J. Martin, The impact of food prices on poverty and food security, Annu. Rev. Resour. Econ. 8 (1) (2016) 329–351.
[2] J. Fanzo, C. Davis, R. McLaren, J. Choufani, The effect of climate change across food systems: implications for nutrition outcomes, Glob. Food Sec. 18 (2018) 12–19.
[3] M.A. Altieri, C.I. Nicholls, A. Henao, M.A. Lana, Agroecology and the design of climate change-resilient farming systems, Agron. Sustain. Dev. 35 (3) (2015) 869–890.
[4] J. Pretty, T.G. Benton, Z.P. Bharucha, L.V. Dicks, C.B. Flora, H.C.J. Godfray, D. Goulson, S. Hartley, N. Lampkin, C. Morris, G. Pierzynski, Global assessment of agricultural system redesign for sustainable intensification, Nat. Sustain. 1 (8) (2018) 441–446.
[5] K. Anderson, S. Nelgen, Agricultural trade distortions during the global financial crisis, Oxf. Rev. Econ. Policy 28 (2) (2012) 235–260.
[6] S.T. Hadebe, A.T. Modi, T. Mabhaudhi, Drought tolerance and water use of cereal crops: a focus on sorghum as a food security crop in sub-Saharan Africa, J. Agron. Crop Sci. 203 (3) (2017) 177–191.
[7] B.N. Aloo, E.R. Mbega, J.B. Tumuhairwe, B.A. Makumba, Advancement and practical applications of rhizobacterial biofertilizers for sustainable crop production in sub-Saharan Africa, Agric. Food Secur. 10 (1) (2021) 57.
[8] T.T. Rubhara, M. Mudhara, O.S. Oduniyi, M.A. Antwi, Impacts of cash crop production on household food security for smallholder farmers: a case of Shamva District, Zimbabwe, Agriculture 10 (5) (2020) 188.
[9] P. Evangelista, N. Young, J. Burnett, How will climate change spatially affect agriculture production in Ethiopia? Case studies of important cereal crops, Clim. Change 119 (2013) 855–873.
[10] M.T. Wakjira, N. Peleg, D. Anghileri, D. Molnar, T. Alamirew, J. Six, P. Molnar, Rainfall seasonality and timing: implications for cereal crop production in Ethiopia, Agric. For. Meteorol. 310 (2021) 108633.
[11] ETHIOPIA T.F.D.R.O. and C.S. AGENCY. Agricultural sample survey. 2021 [cited 2023 April]; Available from: http://www.statsethiopia.gov.et/wp-content/uploads/2021/06/2020_21-2013-E.C-AgSS-Main-Season-Agricultural-Farm-Management-Report.pdf.
[12] P.C. Abhilash, V. Tripathi, S.A. Edrisi, R.K. Dubey, M. Bakshi, P.K. Dubey, H.B. Singh, S.D. Ebbs, Sustainability of crop production from polluted lands, Energy Ecol. Environ. 1 (2016) 54–65.
[13] M. Smith, The application of climatic data for planning and management of sustainable rainfed and irrigated crop production, Agric. For. Meteorol. 103 (1–2) (2000) 99–108.
[14] D. McLaughlin, W. Kinzelbach, Food security and sustainable resource management, Water Resour. Res. 51 (7) (2015) 4966–4985.

[15] I. Lewandowski, M. Härdtlein, M. Kaltschmitt, Sustainable crop production: definition and methodological approach for assessing and implementing sustainability, Crop Sci. 39 (1) (1999) 184–193.

[16] M. Kassie, P. Zikhali, K. Manjur, S. Edwards, Adoption of sustainable agriculture practices: evidence from a semi-arid region of Ethiopia. Natural Resources Forum, Blackwell Publishing Ltd, Oxford, UK, 2009, pp. 189–198. Vol. 33, No. 3.

[17] J.A. Foley, N. Ramankutty, K.A. Brauman, E.S. Cassidy, J.S. Gerber, M. Johnston, N.D. Mueller, C. O'Connell, D.K. Ray, P.C. West, C. Balzer, Solutions for a cultivated planet, Nature 478 (7369) (2011) 337–342.

[18] H.C.J. Godfray, J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, C. Toulmin, Food security: the challenge of feeding 9 billion people, Science 327 (5967) (2010) 812–818 (1979).

[19] R.Q. Grafton, C. Daugbjerg, M.E. Qureshi, Towards food security by 2050, Food Secur. 7 (2015) 179–183.

[20] A.Y. Prosekov, S.A. Ivanova, Food security: the challenge of the present, Geoforum 91 (2018) 73–77.

[21] D.B. Lobell, S.M. Gourdji, The influence of climate change on global crop productivity, Plant Physiol. 160 (4) (2012) 1686–1697.

[22] J. Premanandh, Factors affecting food security and contribution of modern technologies in food sustainability, J. Sci. Food Agric. 91 (15) (2011) 2707–2714.

[23] T.N. Liliane, M.S. Charles, Factors affecting yield of crops, Agron. Clim. Change Food Secur. 9 (2020) 9–24.

[24] K. Heil, P. Heinemann, U. Schmidhalter, Modeling the effects of soil variability, topography, and management on the yield of barley, Front. Environ. Sci. 6 (2018) 146.

[25] H. Jiang, H. Hu, B. Li, Z. Zhang, S. Wang, T. Lin, Understanding the non-stationary relationships between corn yields and meteorology via a spatiotemporally varying coefficient model, Agric. For. Meteorol. 301 (2021) 108340.

[26] S.H. Qader, C.E. Utazi, R. Priyatikanto, P. Najmaddin, E.O. Hama-Ali, N.R. Khwarahm, A.J. Tatem, J. Dash, Exploring the use of Sentinel-2 datasets and environmental variables to model wheat crop yield in smallholder arid and semi-arid farming systems, Sci. Total Environ. 869 (2023) 161716.

[27] A.C. Rencher, G.B. Schaalje, Linear Models in Statistics, John Wiley & Sons, 2008.

[28] S.R. Searle, M.H.J. Gruber, Linear Models, John Wiley & Sons, 2016.

[29] M. Davidian, D.M. Giltinan, Nonlinear models for repeated measurement data: an overview and update, J. Agric. Biol. Environ. Stat. 8 (2003) 387–419.

[30] B.T. West, K.B. Welch, A.T. Galecki, Linear Mixed models: a Practical Guide Using Statistical Software, Chapman and Hall/CRC, 2022.

[31] C.E. McCulloch, S.R. Searle, Generalized, linear, and Mixed Models, John Wiley & Sons, 2004.

[32] S.N. Wood, Generalized Additive models: an Introduction With R, Chapman and Hall/CRC, 2017.

[33] T.J. Hastie, Generalized additive models. Statistical Models in S, 2017, pp. 249–307.

[34] X. Lin, D. Zhang, Inference in generalized additive mixed models by using smoothing splines, J. R. Stat. Soc. B: Stat. Methodol. 61 (2) (1999) 381–400.

[35] Baayen R.H. and Linke M., 2020. An introduction to the generalized additive model. A Practical Handbook Of Corpus Linguistics, pp.563–591.

[36] E.E. Kammann, M.P. Wand, Geoadditive models, J. R. Stat. Soc. C Appl. Stat. 52 (1) (2003) 1–18.

[37] W.R. Tobler, A computer movie simulating urban growth in the detroit region, Econ. Geogr. 46 (sup1) (1970) 234–240.

[38] N. Cressie, Statistics For Spatial Data, John Wiley & Sons, 2015.

[39] M. Blangiardo, M. Cameletti, G. Baio, H. Rue, Spatial and spatio-temporal models with R-INLA, Spat. Spatio Temporal Epidemiol.. 4 (2013) 33–49.

[40] H. Rue, S. Martino, N. Chopin, Approximate bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, J. R. Stat. Soc. B Stat. Methodol. 71 (2) (2009) 319–392.

[41] V. Gómez-Rubio, Bayesian Inference With INLA, Chapman and Hall/CRC, 2020.

[42] F. Lindgren, H. Rue, Bayesian spatial modelling with R-INLA, J. Stat. Softw. 63 (2015) 1–25.

[43] L. Anselin, Spatial Econometrics: Methods and Models, Vol. 4, Springer Science & Business Media, 1988.

[44] Cliff A.D. and Ord J.K., 1981. Spatial processes: models & applications.

[45] A. Getis, J.K. Ord, The analysis of spatial association by use of distance statistics, Geogr. Anal. 24 (3) (1992) 189–206.

[46] L. Anselin, Local indicators of spatial association—LISA, Geogr. Anal. 27 (2) (1995) 93–115.

[47] J.K. Ord, A. Getis, Local spatial autocorrelation statistics: distributional issues and an application, Geogr. Anal. 27 (4) (1995) 286–306.

[48] L. Anselin, Spatial regression analysis in R: a workbook, Urbana 51 (2005) 61801.

[49] G.E. Box, G.C. Tiao, Bayesian Inference in Statistical Analysis, John Wiley & Sons, 2011.

[50] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 1995.

[51] J.M. Bernardo, A.F. Smith, Bayesian Theory, Vol. 405, John Wiley & Sons, 2009.

[52] C.P. Robert, The Bayesian choice: from Decision-Theoretic Foundations to Computational Implementation, Vol. 2, Springer, New York, 2007.

[53] H. Bakka, H. Rue, G.A. Fuglstad, A. Riebler, D. Bolin, J. Illian, E. Krainski, D. Simpson, F. Lindgren, Spatial modeling with R-INLA: a review, Wiley Interdiscip. Rev. Comput. Stat. 10 (6) (2018) e1443.

[54] H. Rue, A. Riebler, S.H. Sørbye, J.B. Illian, D.P. Simpson, F.K. Lindgren, Bayesian computing with INLA: a review, Annu. Rev. Stat. Appl. 4 (1) (2017) 395–421.

[55] T.G. Martins, D. Simpson, F. Lindgren, H. Rue, Bayesian computing with INLA: new features, Comput. Stat. Data Anal. 67 (2013) 68–83.

[56] J. Van Niekerk, H. Bakka, H. Rue, O. Schenk, New frontiers in Bayesian modeling using the INLA package in R, J. Stat. Softw. 100 (2021) 1–28.

[57] Rachmawati R.F.N., 2018. Spatio-temporal models using R-INLA with generalized extreme value distribution in hierarchical Bayes regression. Universitas Pertahanan RI.

[58] R. Bivand, V. Gómez-Rubio, H. Rue, Spatial data analysis with R-INLA with some extensions, J. Stat. Softw. 63 (2015) 1–31.

[59] H. Rue, L. Held, Gaussian Markov random fields: Theory and Applications, Chapman and Hall/CRC, 2005.

[60] J. Pinheiro, D. Bates, Mixed-effects Models in S and S-PLUS, Springer Science & Business Media, 2000.

[61] Y. Mare, D.B. Belay, T. Zewotir, Linear modeling of zonal level crop production in Ethiopia, Heliyon 10 (10) (2024).

[62] Y. Mare, T. Zewotir, D.B. Belay, Additive modeling of zonal level crop production in Ethiopia, Sci. Rep. 15 (1) (2025) 12391.

[63] L. Held, B. Schrödle, H. Rue, Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. Statistical Modelling and Regression structures: Festschrift in Honour of Ludwig Fahrmeir, Physica-Verlag HD, Heidelberg, 2009, pp. 91–110.

[64] L. Pettit, The conditional predictive ordinate for the normal distribution, J. R. Stat. Soc. B (Methodological) 52 (1) (1990) 175–184.

[65] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.

[66] P. Moraga, Geospatial Health data: Modeling and Visualization With R-INLA and Shiny, Chapman and Hall/CRC, 2019.

[67] T.M. Cover, Elements of Information Theory, John Wiley & Sons, 1999.

[68] Hubin A. and Storvik G., 2016. Estimating the marginal likelihood with integrated nested Laplace approximation (INLA). arXiv preprint arXiv:1611.01450.

[69] V. Gómez-Rubio, H. Rue, Markov chain Monte Carlo with the integrated nested Laplace approximation, Stat. Comput. 28 (2018) 1033–1051.

[70] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, A. Van Der Linde, Bayesian measures of model complexity and fit, J. R. Stat. Soc. B (statistical methodology) 64 (4) (2002) 583–639.

[71] S. Watanabe, A widely applicable Bayesian information criterion, J. Mach. Learn. Res. 14 (1) (2013) 867–897.

[72] A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for bayesian models, Stat. Comput. 24 (2014) 997–1016.

[73] M.A. Oliver, R. Webster, A tutorial guide to geostatistics: computing and modelling variograms and kriging, Catena 113 (2014) 56–69 (Amst).

[74] G. Bohling, Introduction to geostatistics and variogram analysis, Kans. Geol. Surv. 1 (10) (2005) 1–20.

[75] A.W. Bowman, R.M. Crujeiras, Inference for variograms, Comput. Stat. Data Anal. 66 (2013) 19–31.

[76] J. Martínez-Minaya, H. Rue, A flexible Bayesian tool for CoDa mixed models: logistic-normal distribution with dirichlet covariance, Stat. Comput. 34 (3) (2024) 116.

[77] A.K. Quaye, C.A. Hall, V.A. Luzadis, Agricultural land use efficiency and food crop production in Ghana, Environ. Dev. Sustain. 12 (2010) 967–983.

[78] K.R. Hakeem, Crop Production and Global Environmental Issues (No. 11748), Springer International Publishing, Switzerland, 2015.

[79] H. Spiertz, Challenges for crop production research in improving land use, productivity and sustainability, Sustainability 5 (4) (2013) 1632–1644.

[80] A.W. Degife, F. Zabel, W. Mauser, Land use scenarios and their effect on potential crop production: the case of Gambella region, Ethiopia, Agriculture 9 (5) (2019) 105.

[81] G.G. Dureti, M.P.J. Tabe-Ojong, E. Owusu-Sekyere, The new normal? Cluster farming and smallholder commercialization in Ethiopia, Agric. Econ. 54 (6) (2023) 900–920.

[82] B. Endalew, A. Elias, K. Yasunobu, Impact of cluster farming on smallholder farmers teff commercialization in Ethiopia, CABI Agric. Biosci. 5 (1) (2024) 18.

[83] G. Tadele, A.A. Gella, A last resort and often not an option at all': farming and young people in Ethiopia, IDS Bull. 43 (6) (2012) 33–43.

[84] T. Guadu, M. Abebaw, Challenges, opportunities and prospects of dairy farming in Ethiopia: a review, World J. Dairy Food Sci. 11 (1) (2016) 01–09.

[85] T. Mihret, F. Mitku, T. Guadu, Dairy farming and its economic importance in Ethiopia: a review, World J. Dairy Food Sci. 12 (1) (2017) 42–51.

[86] B. Minten, Y. Habte, S. Tamru, A. Tesfaye, The transforming dairy sector in Ethiopia, PLoS One 15 (8) (2020) e0237456.

[87] CSA. Area and production for major crops (Private Peasant Holdings, Meher Season) 2019-20 (2012 E.C.). 2020 [cited 2022 March]; Available from: https://www.statsethiopia.gov.et/our-survey-reports/.