Journal of Official Statistics, Vol. 37, No. 1, 2021, pp. 171–211, http://dx.doi.org/10.2478/JOS-2021-0008

Measuring and Modeling Food Losses

Marco Mingione[1], *Carola Fabi*[2], *and Giovanna Jona Lasinio*[1]

Within the context of Sustainable Development Goals, progress towards Target 12.3 can be measured and monitored with the Food Loss Index. A major challenge is the lack of data, which dictated many methodology decisions. Therefore, the objective of this work is to present a possible improvement to the modeling approach used by the Food and Agricultural Organization in estimating the annual percentage of food losses by country and commodity. Our proposal combines robust statistical techniques with the strict adherence to the rules of the official statistics. In particular, the case study focuses on cereal crops, which currently have the highest (yet incomplete) data coverage and allow for more ambitious modeling choices. Cereal data is available in 66 countries and 14 different cereal commodities from 1991 to 2014. We use the annual food loss as response variable, expressed as percentage over production, by country and cereal commodity. The estimation work is twofold: it aims at selecting the most important factors explaining losses worldwide, comparing two Bayesian model selection approaches, and then at predicting losses with a Beta regression model in a fully Bayesian framework.

Key words: Bayesian variable selection; Beta mixed model; SDG 12.3.

1. Introduction

The Transforming our world: 2030 Agenda for Sustainable Development (2015), approved by all the Member States of the United Nations (UN) in September 2015, officially came into force on 1st January 2016. The Agenda includes 17 *Sustainable Development Goals* (SDGs) and 169 *Targets* supported by a global monitoring framework with 231 *Indicators*, which were established to track progress. Some of these goals are groundbreaking: indeed, for many of them, there was no specific indicator, methodology, nor underlying data for the measuring, and food losses reduction fell into this category. More precisely, the indicator to measure and monitor food losses, associated with Target 12.3, was initially classified in Tier III, meaning that an indicator and data collection method needed to be specifically developed for that purpose. One significant challenge is the lack of reliable estimates of the level of losses (and waste) worldwide, particularly in developing countries, for numerous reasons (Fabi et al. 2018). Preliminary work indicates that food losses and waste remain unacceptably high, impacting economic efficiency and natural resource usage, and contributing to inefficient food systems. The widely quoted advocacy study "*The Global Food Loss and Waste – extent, causes, and prevention*" (Gustavsson et al. 2011), published

[1] University of Rome La Sapienza, Department of Statistical Sciences, Piazzale Aldo Moro 5, Rome, 00185, Italy. Emails: marco.mingione@uniroma1.it and giovanna.jonalasinio@uniroma1.it
[2] Food and Agriculture Organization of the United Nations, Statistics Division, Rome, Italy. Email: carola.fabi@fao.org
Acknowledgments: The views expressed in this publication are those of the authors and do not necessarily reflect the views or policies the Food and Agriculture Organization of the United Nations (FAO).

by the Food and Agricultural Organization of the United Nations (FAO), estimates that yearly global food loss and waste account for 30% of the overall production, which is equivalent to almost USD 1 trillion. Recently, model-based estimates in the *State of Food and Agricultural* (SOFA) Report (FAO 2019) confirmed that food losses on the supply side alone (after harvest and up to but excluding the retail level) amounted to almost 14% of agriculture production and were worth at least USD 400 billion in 2016.

1.1. Literature on the Measurement and Estimation of Global Food Losses

Food loss measuring and monitoring is not a novel issue among experts in both the private and public sectors. The UN General Assembly addressed the problem back in 1975 and passed a resolution calling for "a 50% reduction of post-harvest losses by 1985". In 1976, FAO identified the major constraints causing post-harvest losses focusing on staple crops, including grains and pulses. Two years later, the FAO produced an action program that led to the development of a standard terminology and suitable methodology for the measurement of post-harvest losses, formalized in the milestone publication, "*Postharvest Grain Loss Assessment Methods*" (Harris and Lindblad 1978). Some methods and techniques explained in the manual were later revised by Boxall (1986) over the period 1980–1986, in an attempt to simplify them. Additionally, in 1980 FAO published guidelines for the *Assessment and Collection of Data on Post-Harvest Food-Grain Losses* (FAO 1980) to support the implementation of a statistical methodology combining objective measurements with statistical survey sampling techniques to collect data and produce accurate survey-based post-harvest loss estimates. Many other studies followed these efforts in modeling and estimating losses. The most relevant and recent ones that have received the highest level of worldwide consensus are: *The African Post-harvest Losses Information System* (APHLIS), which developed a calculator to estimate cumulative post-harvest losses over the entire value chain, as a percentage of production for nine cereals in Sub-Saharan African countries (SSA); *The Global Food Loss and Waste – extent, cause, and prevention* report (Gustavsson et al. 2011), that changed the world perception on food loss and waste and uses a mass balance approach to quantify the volumes of food loss and waste at the global level; *Imputation of Loss Ratios,* a technical report by an FAO consultant, Klaus Grünberger (2013), who developed an econometric model to estimate loss using causal factors and covariates, such as countries' infrastructure, national income level, geographic region, and commodity groups. The causal factors were not significant, hence the model was abandoned.

All these efforts have been hindered by little available data, which reflects the low priority given to post-harvest losses until recently, and to the objective complexity and cost of food loss data collection. These constraints persist and affect the quality of the estimates and, consequently, the reliability of results. The dire lack of data, an international definition of food losses and a recognized methodology to monitor loss reduction underpinned the need to develop a standardized approach for measuring, collecting data, and modeling food losses. A comprehensive methodology including a measurable definition, an indicator, an aggregation method, an estimation model and a range of data collection methods and tools has been developed by FAO to help countries measure food losses and monitor progress against SDG target 12.3 (FAO 2019).

This article aims to present an improved model capable of estimating food losses at the country-commodity level. The new model considers a set of explanatory variables that scientific literature has consistently identified as the causes or proxies of causes of losses in all countries of the world. The purpose of using explanatory variables is to link losses with their causal factors at the country-commodity level to support decisions on interventions, investments, and policy-making. Our model's main feature is that it builds on the finding of previous efforts and works toward overcoming their weaknesses.

1.2. Definitions

In recent years, Food Loss and Waste (FLW) became a priority issue on the global agenda, for both the public and private sector, as one aspect of sustainable global food systems. In the absence of a commonly agreed definition, the various stakeholders have developed their definitions of food loss and waste, albeit a pre-condition for a harmonized methodological approach and data comparability is agreement on the terminology. For this reason FAO, under the aegis of the Save Food initiative, undertook the development of a *FLW Definitional Framework* in consultation with national and international stakeholders, building on the previous definitions found in the literature and laying the foundation for a consistent methodology. In what follows, we will only report the most important definitions required for a proper comprehension of this work. For further details, we highly recommend to take a look at the whole document (FAO 2014). In particular, the main definitions include:

- **Food Supply Chain (FSC):** the connected series of activities to produce, process, distribute, and consume food; and
- **Food Loss:** the decrease in quantity or quality of food.

For the sake of measurability and consistency with the SDG 12.3 target formulation, an operational definition of "Food Loss" was added to the *Definitional Framework* in 2016 (unpublished) drawn from FAO's annual questionnaire on agriculture production whereby:

- **Food Losses** are crop and livestock product losses and cover all quantity losses along the supply chain for all utilizations (food, feed, seed, industrial, other) up to, but not including, the retail/consumption level. Losses of the commodity as a whole (including edible and non-edible parts) and losses direct or indirect, which occur during storage, transportation, and processing, also of relevant imported quantities, are therefore all included; and
- **Food Waste** occurs from retail to the final consumption/demand stages.

1.3. SDG Framework

The UN has defined sustainable development as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs". At the same time, this means promoting resource and energy efficiency, sustainable infrastructures, providing access to basic services, green and decent jobs, and a better quality of life for all. In this respect, one aspect is to promote a "Responsible Consumption and Production" (SDG 12). The third target under this goal (Target 12.3) states "By 2030, [to] halve per capita global food waste at the retail and consumer levels and reduce food losses along production and supply chains, including post-harvest losses". FAO and UN Environment are the custodians of SDG 12.3 and, for consistency with policy objectives, relevance, and measurement, the indicator has been split into two distinct sub-indicators that focus on losses on the supply side and waste on the demand and consumption side of the food systems, respectively. This work only focuses on the first sub-indicator, namely the one on losses, which will be introduced in Subsection 1.4. For a complete overview of the SDG and related Targets, the authors refer to *Global indicator framework for the Sustainable Development Goals and Targets of the 2030 Agenda for Sustainable Development* which is accessible at https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202020%20review_Eng.pdf.

1.4. The Food Loss Index

At the global level, the GFLI, *Global Food Loss Index* (FAO 2019) is a composite indicator, built as a weighted average of countries' FLIs, Food Loss Indices (FAO 2019). A country FLI is a fixed-base index that aggregates the losses of ten key commodities in the five main food groups using economic weights (value of production in the base period). FAO is partnering with national and international stakeholders to foster data collection along the supply chains and build the evidence base for these commodities. Although the FLI uses aggregated percentage losses along the supply chain, more disaggregated data at different stages of the value chain (e.g., farm, transportation, storage, processing, and wholesale) is needed to decide on appropriate interventions. The countries' FLIs summarise complexities of food loss and their dynamics to provide decision-makers with an overview of the magnitude of the problem at the national level and an overall monitoring indicator.

The food loss data set can be treated as a longitudinal data set for a multivariate outcome across different countries. In the ensuing sections, we will refer to the observed (or estimated) loss percentage for country *i,* commodity *j* at year *t* as l_{ijt}. Therefore, the FLI for country *i* in year *t* is defined as:

$$FLI_{it} = \frac{\sum_j l_{ijt} \cdot q_{ijt_0} \cdot p_{jt_0}}{\sum_j l_{ijt_0} \cdot q_{ijt_0} \cdot p_{jt_0}} \cdot 100 \tag{1}$$

where t_0 is the reference year; q_{ijt_0} the production quantities by country and commodity in the reference year, available in FAO's corporate statistical database (FAOSTAT 2016); p_{jt_0} the fixed price (in USD) set by commodity for the (t_0−1)-(t_0 + 1) average. At present, the reference year is set to 2015 (the year in which countries adopted the SDGs), while l_{ijt} can be either survey-based or model-based. The food loss percentages at the commodity or country level can be interpreted as the average percentage of supply that does not reach the retail stage. The FLI shows the relative change in percentage food loss for country i over time, compared to the base year. Finally, using weights proportional to the total value of agricultural production in the base year, the country indices can be aggregated to build the GFLI. To achieve SDG 12.3, both GFLI and FLIs should ideally show that post-harvest losses decrease compared to the base period from a base value of 100.

1.5. Basic Data Constraints

Primary data on losses are seldom compiled within the national statistical systems worldwide: only 39 countries out of 185 reported losses for one product or more in FAO's annual Questionnaires on Agricultural Production, including a section on product utilization. Moreover, reporting on losses has increased slightly in recent years and data was even more scarce in the past period. Data on utilization, including losses, stock changes, and food supply, is used to compile the Supply Utilization Accounts (SUA) and Food Balance Sheets (FBS). The FBS framework defines agricultural production net of harvest losses and collects loss estimates net of harvest losses. It is worth noting that only 7% of loss data in FAOSTAT's FBS domain (FAOSTAT 2016) was officially reported by the countries in the period 1990–2016. The remaining 93% of records are estimated or considered null. In conclusion, since representative data on losses are very scarce, the FLI will be model-based. However, with the strong emphasis put on SDG 12.3 and the need for evidence-based policy-making, one has to expect an increase in data availability in the future years. The methodology provided herein attempts to further refine the 2016–18 FAO developed model, described in Subsection. 1.6.

1.6. FAO Current Modeling Approach – SOFA 2019

FAO developed a random effects model able to exploit panel data information, that is in a cross-section – by commodity and country – and longitudinally over time to estimate missing loss data and compile the FLI of all countries (FAO 2019). The model is part of the methodology for monitoring progress against SDG 12.3 (see SDG indicators metadata at https://unstats.un.org/sdgs/metadata/). Results were first published in the State of Food and Agriculture 2019 edition (FAO 2019) and stated that global food losses along the supply chain, up to but excluding the retail level, are almost 14% of 2016 total production. At present, FAO can disseminate loss estimates at the global, regional, and commodity group level. The model supplements the 7% officially reported loss data along the supply chain with two additional data sets. The first one is a data set of food losses built from a literature review to increase the coverage. The second one is a data set composed of over 200 possible explanatory variables from various international sources (International Energy Association, World Bank, FAO, and more), possibly representing causal factors or proxy variables for the causes of losses. These causal factors can be grouped under common categories to be easily managed by a model. These categories are Energy, Inputs and Associated Costs; Investment and Monetary Policy; Social and Economic Factors; Storage, Transportation, and Logistics; Weather and Crop Cycles. The random forest algorithm was used to standardize variables' selection and choose the five most important ones by commodity grouping. The purpose was to better capture the variation in the causes of losses by country or region and commodity. Where the observations by country and commodity are fewer than three, a bare minimum to run the model for a country-commodity combination, available information has been clustered by commodity group on the assumption that causes and rates of losses are more similar within the groups than across them (for example losses of maize and lentils are more similar than losses of maize and fresh milk). The same assumption applies to the value chain (e.g., traditional, capital-intensive, vertically integrated, and more) and solutions (improved farm practices, infrastructures, cool chain, and others). Clustering scarce data evened out the impact of outliers on the results. The coexistence of country-level estimates and cluster-level estimates required a model hierarchy to fill in the results matrix. All the methodological choices have been dictated by the need to overcome data scarcity.

The rest of the article is structured as follows: Section 2 describes the available data and some preliminary results on data consistency; Section 3 delves into the methodology. In Section 4, we report and analyze the model results. Section 5 is dedicated to the concluding remarks and discussion.

2. Available Data

We worked with official loss data extracted from FAOSTAT (2016). The data set covered 138 countries and 145 different commodities, with the most extended time series starting in 1961 and ending in 2015, and included a total number of 18,472 records. A preliminary exploratory analysis of the data highlighted some critical issues. First, some country-commodity combinations presented loss levels larger than their total production: these losses characterize several import-dependent countries in which domestic supply consists mainly of imported produce. The FLI methodology deals with import-dependent countries by changing the denominator in the ratio. In this work, we did not introduce any exception and therefore we excluded these records from the analysis. Second, 4264 of country-commodity-year combination records were equal to zero. Zero losses on such a large scale are unlikely and instead point at under-coverage or missing data interpreted as nil amounts. Moreover, the comparison of FAOSTAT data to loss factors found in the literature showed a systematic difference. The SUA seems to underestimate the actual losses within the countries and the explanation is manifold (Fabi et al. 2018). Case studies in the literature tend to focus on countries where losses are high, and the problem is more acute, representing an upper boundary. On the contrary, nation-wide estimates average losses across all value chains, including the more efficient ones. Also, losses are sometimes obtained as the balance for quantities that cannot be accounted for in the SUA. Therefore, SUA data can be considered the lower boundary. Indeed, FAOSTAT data showed a global loss average of 7.2% over the whole data set, which is 9.4% when excluding zero values. Another data constraint and challenge to the modeling framework is that countries tend to use carry-forward estimates on loss percentages, on the grounds that systemic losses do not change quickly over time, which at the same time removes any trend from the time series (see appendix, Subsection. 6.1, Figure 7a and 7b for clarification). Additional information was gathered from more than 300 publications and reports from various sources to increase observations and reduce the noise in the data. These sources included reports from international organizations, such as the World Bank, GIZ (Gesellschaft fur Internationale Zusammenarbeit), FAO, IFPRI (International Food Policy Research Institute), sub-national reports, and academic institutions (Fabi et al. 2018). All data have been consolidated in a database that is continuously updated and accessible at http://www.fao.org/platform-food-loss-waste/flw-data/en/.

2.1. Cereals and Cereal Products

This food category includes the largest share of available loss data. Hence, we focused our modeling efforts on cereals.

More precisely, cereals data include 66 countries and 14 different commodities, amounting to 196 country-crop combinations ($< 66 \times 14$ as not all countries produce all cereals). A simple average of the available data gives a loss percentage of 5.6%, but there is variability both by country and commodity (for further details see appendix, Subsection 6.1, Figures 8 and 9). Figure 1 represents a clear snapshot of the data availability: each square identifies a country-crop combination, and it is colored according to the country-crop temporal average.

The majority of estimates are provided by countries in North America and Europe (NAE). In particular, 111 out of the 196 country-crop combinations (more than 50%) come from NAE, whose average loss percentage is the lowest (only 3.24%), as reported in Table 1. Sub-Saharan Africa (SSA) records the highest loss percentages, with losses amounting to 24% of total production, although only five SSA countries reported data on losses. This unbalanced data distribution does not introduce any bias in our methodology because the estimation is carried out at the country-crop level. However, some bias may be introduced when losses are aggregated at the global level. In this case, the weights used to calculate the GFLI should be proportional to each country's agricultural sector size. Nevertheless, the representativeness of single countries or single macro-regions can differ significantly. We do not want to compile the global losses by estimating missing country data exclusively from countries within the same region (e.g., losses in Asian countries estimated only losses of the Asian countries). Few available countries will determine the estimates too heavily. If the few countries are not representative of the region, the regional and ultimately the global estimates will underestimate/overestimate the actual loss level.

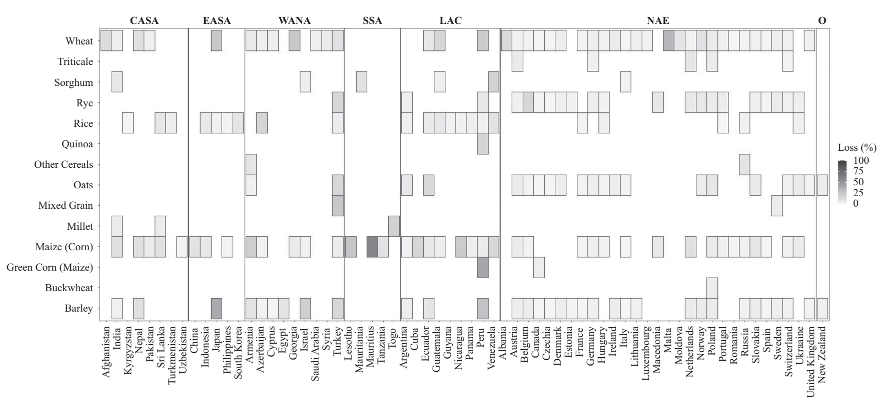


Fig. 1. Available data for cereals. Each square represents the temporal average loss by country and commodity. Countries have been ordered on the x-axis with respect to the SDG region they belong Central Asia and Southern Asia (CASA), Eastern Asia and South-eastern Asia (EASA), Western Asia and Northern Africa (WANA), Sub-Saharan Africa (SSA), Latin America and the Caribbean (LAC), North America and Europe (NAE), Oceania (O).

Table 1. Average loss (%) and number of distinct country-crop combination by SDG region: Central Asia and Southern Asia (CASA), Eastern Asia and South-eastern Asia (EASA), Western Asia and Northern Africa (WANA), Sub-Saharan Africa (SSA), Latin America and Caribbean (LAC), North America and Europe (NAE), Oceania (O).

	CASA	EASA	WANA	SSA	LAC	NAE	O
Loss(%)	5.86	10.53	8.58	24.09	9.67	3.24	3.27
nObs	16	9	24	5	29	111	2

3. Methodology

This section will describe our proposal. We will first define the steps in our statistical protocol and then delve into the model in detail. Our model will estimate losses by country, commodity, and year for cereals in a full Bayesian framework, so as to reduce the impact of critical issues in the data described in Section 2. After dealing with missing data in the available predictors, we build a Beta regression model with a latent component (Wu 2009). The latent component captures variations at the country-crop level due to missing information on other known causal factors such as the time of harvest, rainfall on crop areas, and other variables that should be measured at crop level.

3.1. Missing Predictors Imputation

A total number of $K = 34$ explanatory variables were considered in the loss estimation model. Most variables are proxies for relevant explanatory factors commonly found in the scientific literature. Similar to the loss imputation model of the SDG methodology, these factors can be grouped into categories relating to *Energy*, *Economic Factors*, *Transportation and Logistics*, *Building Materials* and *Weather and Crop Cycles* (see appendix, Subsection 6.2). However, not all the variables are available for all countries and years, highlighting a severe missing data issue as the *missingness* in the set of predictors is almost 19%. In particular, 16 out of the 34 variables contain at least one missing value and 13 out of these 16 have more than 30% of missing values overall. Assuming a MAR mechanism, we consider three non-parametric missing value imputation methods: the missForest algorithm (Stekhoven and Bühlmann 2012), Multiple Imputation by Chained Equations (MICE) approach (White et al. 2011) and k-nearest neighbours (K-NN) as in Franzin et al. (2016). With each imputed data set, we estimate the model in Equation (5) and compare results in terms of variable selection and prediction accuracy. We do not report all the details, but we simply note that the three imputation methods produce similar outputs in terms of final model performances. We decided to keep the imputed data set with missForest for its flexibility with respect to assumptions on data collection and distribution. In the seminal paper (Stekhoven and Buhlmann 2012), the authors show how missForest generally outperforms the two other imputation methods. Also, as demonstrated empirically not only with our set of data (Waljee et al. 2013; Cihan 2018), the missForest algorithm yields better results, especially in terms of out of sample prediction error. This happens because of its non-parametric nature, which allows for the imputation of mixed-type data. Being based on a Random Forest algorithm (Breiman 2001), it has no need for tuning parameters, nor does it require, any assumptions about the distributional aspects of the data. Besides, it offers a way to assess the quality of an imputation without the need for setting aside test data nor performing cross-validations. In particular, the full potential of missForest is deployed when the data include complex interactions or non-linear relations between variables of unequal scales, as it is in our case study.

3.2. Beta Regression

Food losses are expressed as percentage of the total production, hence the Beta distribution is the most natural assumption for their modeling. Indeed, the class of Beta regression models, firstly proposed by Ferrari and Cribari-Neto (2004), is commonly used to model random variables that take values in the open standard unit interval (0, 1). The main assumption is that the dependent variable is Beta-distributed and its mean $\mu \in (0, 1)$ is related to a set of regressors through a linear predictor with unknown coefficients and a link function $g : (0, 1) \rightarrow \mathbb{R}$, strictly increasing and twice differentiable. The model also includes a precision parameter $\phi > 0$ (independent from μ), which may be constant or may depend on a set of predictors through a link function as well. This approach has the advantage of naturally incorporating features such as heteroskedasticity or skewness, which are commonly observed in data taking values in the standard unit interval.

We assume that our outcome variable $y_i, \ldots, y_n$ is the realization of a random sample such that $y_i \sim Beta(\mu_i, \phi)$, $i = 1,...,n$. Then, the Beta regression model is defined as $g(\mu_i) = x_i'\beta$; where β is the $k \times 1$ vector of unknown parameters and x_i is the vector of k predictors. The logit function represents the most common choice as link function for $g(\cdot)$ due to its shape and ease of interpretation, such as in any typical *Generalized Linear Model* (GLM) framework.

3.3. Bayesian Variable Selection

Given the high dimensionality of the problem, which counts 34 predictors, our first objective is to find a robust selection method for the most relevant factors that can explain losses' behavior. The goal is to find the subset of variables that can simultaneously catch the dependencies and dynamics driving food losses, but that are also meaningful for policy making. This issue is of paramount importance and raises several challenges. A known problem when the number of relevant variables is large, is to account for possible collinearity in order to avoid conflicting results when assessing the importance of strongly correlated predictors (Ijarchelo et al. 2016).

In a Bayesian perspective, variable selection falls in the more general framework of *model choice* and can be addressed with various possible approaches. In this work, we consider two main alternative solutions: the spike and slab technique (Mitchell and Beauchamp 1988) and the horseshoe prior (within the class of shrinkage priors), introduced by Carvalho et al. (2010). Both techniques are based on specific choices for the prior distributions associated to model's coefficients. The first approach belongs to the more general class of discrete mixtures, initially discussed in Mitchell and Beauchamp (1988) and George and McCulloch (1993). It models prior knowledge on coefficients s with a prior comprising both a point mass at zero and an absolutely continuous alternative; the second approach, introduced by Tibshirani (1996), models fis prior distribution with absolutely continuous *shrinkage priors* centered at zero.

3.1.1. Spike and Slab

Spike and slab is considered as the gold standard to combine variable selection with the estimation of the regression parameters. With this technique, variables are chosen by estimating the *posterior* probability of all the models within the considered class (O'Hara et al. 2009), based on the *a priori* knowledge or expectation that only few variables truly impact on the outcome. The main assumption is that the prior distribution of the k-th regression parameter is a mixture of two components: a probability mass either exactly at or around zero (spike) and a flat distribution (slab) elsewhere. Therefore, this prior is often written as:

$$\beta_k|\gamma_k, c, \epsilon \sim \gamma_k N(0, c^2) + (1-\gamma_k)N(0, \epsilon^2), \tag{2}$$

where $\epsilon \ll c$ and where $\gamma_k \in \{0, 1\}$, denoting absence or presence of the k-th variable in the model. If ϵ is set to 0, then the spike is taken to be a Dirac δ_0, at the origin.

In Kuo and Mallick (1998), γ_k is embedded in the regression equation as follows:

$$y_i = \sum_{k=1}^{K} \beta_k \gamma_k x_{ik} + \epsilon_i. \tag{3}$$

Independent priors are typically assumed for β_k, γ_k and the response variance. In particular, $\gamma_k \sim Ber(p_k)$, namely a Bernoulli distribution with success probability p_k that reflects the preference for including the k-th predictor in the model building: for example, $p_k = 0.5, \forall k$ is associated to prior belief of the equally likely relevance of all possible 2^k sub-models.

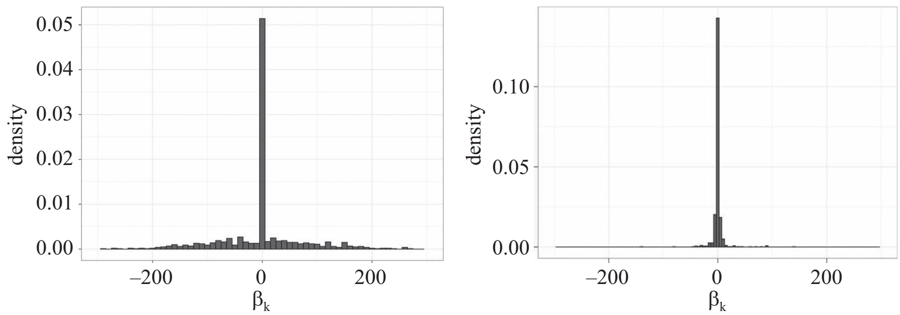
Once the model has been set up, it is usually fitted using Markov Chain Monte Carlo (MCMC) and the variable selection part of the model entails estimating γ_k. As a result, the posterior inclusion probability $\mathbb{E}\left[\gamma_k|\mathrm{y}\right]$ can easily be calculated as the mean value of the indicator γ_k as follows: $\mathbb{E}\left[\gamma_k|\mathrm{y}\right] = P(\gamma_\mathrm{k} = 1|\mathrm{y}) = \frac{1}{T}\Sigma_{t=1}^{T}\gamma_\mathrm{k}^{(t)}$ where T denotes the total number of posterior samples. The selection rule consists merely of keeping those variables with posterior inclusion probability larger than a given threshold. If the threshold is set at 0.5, then the selection criterion is known as the *Median Probability Model* (MPM) by Barbieri et al. (2004). This criterion is known to be robust as it is the optimal predictive model under a squared error loss function with certain regularity conditions and the selected variables appear in at least half of the visited models (Barbieri et al. 2004). The orthogonality of the design matrix is required in all the sub-model scenarios to satisfy these conditions. If this is not the case, inference based on marginal inclusion probability could be incorrect.

3.3.2. Horseshoe Prior

This approach assumes that each coefficient β_k, is *a priori* distributed as a scale mixtures of Normal distributions:

$$\begin{aligned} &\beta_k|\lambda_k, \eta \sim N(0, \lambda_k^2\eta^2), \\ &\lambda_k \sim C_+(0, 1), \quad \eta \sim C_+(0, 1) \end{aligned} \tag{4}$$

where C_+(0,1) represents the half-Cauchy distribution, λ_k is called *local* shrinkage parameter and η is the *global* shrinkage parameter (Carvalho et al. 2009). The horseshoe is named after the shape of the *shrinkage coefficient,* which is $\frac{1}{(1+\lambda_k^2)} \sim Beta(\frac{1}{2}, \frac{1}{2})$ and can be interpreted as the posterior amount of weight that the posterior mean of β_k places on 0. Horseshoe prior's main advantage lies in its flat tails allowing for strong signals to remain large *a posteriori* and in its infinitely tall spike at the origin that severely shrink the β_ks that are very likely to be zero. It can be easily noticed that setting $\epsilon = 0$ in Equation (2), generates a prior distribution very close to Equation (4) that allows for only two values, that is 0 and 1, instead of assigning continuous priors to γ_k, as in the case of the horseshoe. For the sake of clarity, Figure 2 shows the distribution of the prior on β_k in the case of the spike and slab (2a) and the horseshoe (2b).



(a) (b)

Fig. 2. Prior distribution of the k-th regression coefficient in the case of spike and slab prior (a) and horseshoe (b).

3.4. The Model – Our Proposal

Our two alternative proposals combine the procedure described in Subsection 3.2 with the ones described in Subsubsections 3.3.1 and 3.3.2, respectively. Let l_{ijt} be the observed loss percentage in country i, for cereal j at year t and x_{ikt} be the value of the k-th explanatory variable for country i at time t. The model with variable selection is expressed as follows:

$$l_{ijt} \sim Beta(\mu_{ijt,}\phi), \forall i,j,t$$

$$logit(\mu_{ijt}) = \beta_i t + \sum_k x_{ikt}\beta_k^* + v_{ij}, \quad \forall i,j,t \tag{5}$$

$$v_{ij} \sim N(0,\tau^2), \quad \phi \sim Unif(5,150), \quad \tau \sim Gamma(4,0.1),$$

where $\beta_i t$ is a temporal linear trend specific to country i and v_{ij} is the latent component describing the nested *commodity within country* effect. We consider different trend parameters for each country to capture different general behaviors dictated by country-specific policies or climate conditions, or other unobserved factors. The temporal trend was always included to detect generally well- or poor-performing countries in terms of the FLI. Parameter ϕ is known as the precision parameter, since for a given μ_{ijt}, a larger ϕ implies a smaller variance for l_{ijt}. We also adopted a constant precision τ^2 across countries and commodities after estimating several models with different precision parameters (e.g., country-specific, crop-specific, or their sum) that did not yield significantly different estimates. Finally, according to the prior distribution ascribed to the regression coefficients β_k^* we can obtain either the spike and slab model (3) or the horseshoe model (4). While hyperparameters for the shrinkage priors represent standard statistical choices commonly used in the Bayesian variable selection procedures, hyperparameters for ϕ and τ are set to obtain weekly informative priors. In the spike and slab model, we set $\gamma_k \sim Ber(p_k)$, adding a further level to the model by treating p_k with a $\beta(5,5)$ so that all the models were equally likely to be selected *a priori*. The prior distribution on the trend coefficients β_i is $N(0,1000)$.

3.5. Watanabe-Akaike Information Criterion

We compared the performances of the two variable selection procedures using the *Watanabe-Akaike Information Criterion* (WAIC) proposed by Watanabe (2010). The main assumption, which should hold in our model setting, is that the observed values are conditionally independent given the parameters. If the model fits our data, then parameter estimation should minimize the expected log-pointwise predictive density. More precisely, let $lpd = \sum_{i=1}^{n} log \int p(y_i|y)p(\boldsymbol{\theta}|y)d\theta$ is the log-pointwise predictive density and $p = \Sigma_{i=1}^{n} V_{post}[log(p(y_i|\boldsymbol{\theta}))]$ is the estimated *effective number of parameters,* that is, the sum of the posterior variance (V_{post}) of the log-predictive density for each data point. Following Vehtari et al. (2017), the expected log-pointwise predictive density is given by $elpd = lpd - p$. The WAIC is then obtained as $WAIC = -2 \cdot elpd$. We use the WAIC instead of the *Deviance Information Criterion* (DIC) for two reasons: (1) WAIC has the desirable property of averaging over the posterior distribution rather than conditioning on a point estimate, which is particularly relevant in a predictive context (Gelman et al. 2014); and (2) the DIC has a weaker theoretical justification (Celeux et al. 2006; Spiegelhalter et al. 2014). Furthermore, since the final objective is to predict food losses, the WAIC is the more appropriate choice as it is asymptotically equivalent to the Bayesian leave-one-out cross-validation (Watanabe 2010) and hence it can be seen as a measure of a model's predictive performance.

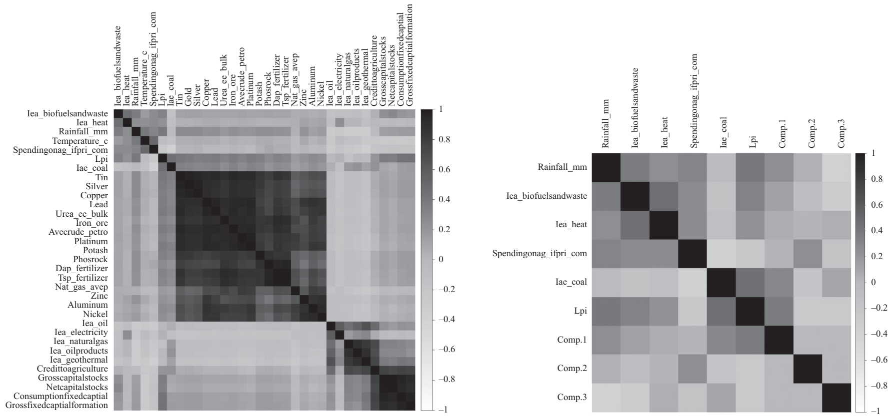
4. Results

In this section, we present the results of our modeling effort. Subsection 4.1 will cover data pre-processing, including dimensionality reduction; results of the variable selection are reported in Subsection 4.2, and the out-of-sample predictions in Subsection 4.3. Note that all the variables were standardized before the quantitative analysis.

4.1. Dealing With Collinearity

Following Subsection 3.3, we checked for the presence of multicollinearity in the predictors. We notice that the estimation of the posterior inclusion probabilities in the spike and slab framework is not reliable in the presence of severe collinearity. In particular, following Bhadra et al. (2019a), optimality can be achieved in terms of parameters' estimation if the design matrix is well-conditioned (e.g., orthogonal). The design matrix orthogonality ensures that no information is shared among the predictors, while collinearity has the effect of blurring distinctions between predictors in the variable selection process.

To this purpose, we first computed the correlation matrix (Figure 3a). Three different groups of strongly correlated variables can easily be pointed out: the first one (the big black square at the center of Figure 3a) includes all metals' prices (e.g., potash, silver, iron, gold, lead, etc.); the second one includes the prices of electricity, natural gas oil and derived products provided by the International Energy Agency (IEA); the third group (bottom right corner) includes all the economic variables from national accounts (such as net capital stocks) and credit to agriculture. We carried out a preliminary dimensionality reduction on the predictors using a simple principal component analysis (PCA) on 27 out of the 34 standardized variables in the three groups leaving out the seven variables (i.e., rainfall (mm), temperature (C), biofuels, heat, coal, LPI, spending on agriculture) in the top left square of Figure 3a, as they are not highly correlated with the others or among them. Results show that three components can explain 81% of the total variance. The first (principal) component can be interpreted as a proxy for input prices with metal prices (iron, silver, copper, etc.) for implements and infrastructure, and fertilizers' prices (potash, urea, etc.) for growing crops (see Figure 10a in the appendix, Subsection 6.3). The second component can



(a) (b)

Fig. 3. Graphical representation of the correlation matrices: (a) before dimensionality reduction of 27 variables out of 34, (b) after dimensionality reduction.

be interpreted as a proxy for investment in agriculture (capital stocks, credit to agriculture, and more), but with negative loadings (see Figure 10b in the appendix, Subsection 6.3). In other words, lower values of this component correspond to higher values of the considered variables. The third component is a proxy for energy's price (oil, natural gas, electricity, and more, see Figure 10c in the appendix, Subsection 6.3). The final data set includes ten almost completely uncorrelated variables (three components plus seven standardized variables), whose correlation matrix is shown in Figure 3b.

4.2. *Estimation*

The estimation was carried out using JAGS (Plummer 2003), a well-known software for Bayesian model estimation, which uses Gibbs Sampler and the Metropolis Within Gibbs Sampler algorithms. For each variable selection approach, we ran the MCMC algorithm with two chains, 120,000 iterations, a burn-in of 60,000 iterations, and a thinning of ten, keeping 6,000 samples from each chain for inferential purposes. Coding examples for both the estimation and prediction of the model with the different prior settings are available in appendix, Subsection 6.4. From now on, we will refer to the model with spike and slab priors as **M1** and to the model with horseshoe prior as **M2**.

4.2.1. WAIC

The WAIC is equal to -19510.6 for M1 and -19522.8 for M2, meaning that the two selection approaches are substantially equivalent, with M2 performing slightly better in terms of goodness of fit.

4.2.2. Posterior Inclusion Probabilities and Selected Variables

Recall that the spike and slab procedure allows for estimating the posterior inclusion probability for each predictor. We chose the *Median Probability Model* as the selection rule (see Subsubsection 3.3.1), hence we kept all the variables with posterior inclusion probability larger than 0.5.

The horseshoe priors do not provide a straightforward variable selection technique, hence we decided to keep all variables with coefficients whose 95% posterior credible intervals did not include the 0 value. In other words, the whole point of shrinking priors is to shrink to zero coefficients that are not significant, according to the classical likelihood definition. Figure 4 illustrates the outcome of the selection step. Both techniques select four variables, that is, biofuels (price), spending on agriculture (i.e., the agriculture share in GDP, which is a proxy for the agricultural sector relative importance in the national economies), the second principal component (comp. 2), and the third principal component (comp. 3). Both the computed point estimates and the credible intervals are comparable. In particular, according to M1, biofuels, spending on agriculture, and comp. 2 are included in the model with probability equal to 1, and comp. 3 is included with probability equal to 0.56 (see Table 3 in the Appendix, Subsection 6.5, for the exact numerical results). The temperature is selected only by M2, while according to M1, its posterior inclusion probability $\hat{\gamma}_k$ is equal to $0.0075 \simeq 0$.

Biofuels has the most considerable effect (in absolute value) on the outcome with a positive coefficient. This variable, measured by the IEA, represents solid biofuels, liquid biofuels, and biogases produced with industrial and municipal waste. Biofuels are a possible utilization of cereals, both the full grains and its waste or discarded quantities. In this respect, a commissioned study by FAO (Kuiper and Cui 2020) found that reducing food losses could decrease agricultural prices, which would benefit the production of meat and biofuels, through lower agricultural input prices. This study can help explain the correlation between biofuels price anpricd losses. An increase in biofuels price would increase the demand for input crops and absorb larger amounts of cereals for industrial uses, thus reducing the quantities ultimately lost. Unfortunately, biofuel data are often based on small sample surveys or other incomplete information. Thus, the data give only a broad overview of the biofuel sector and are not strictly comparable across countries (IEA 2019). Spending on agriculture has the second-largest

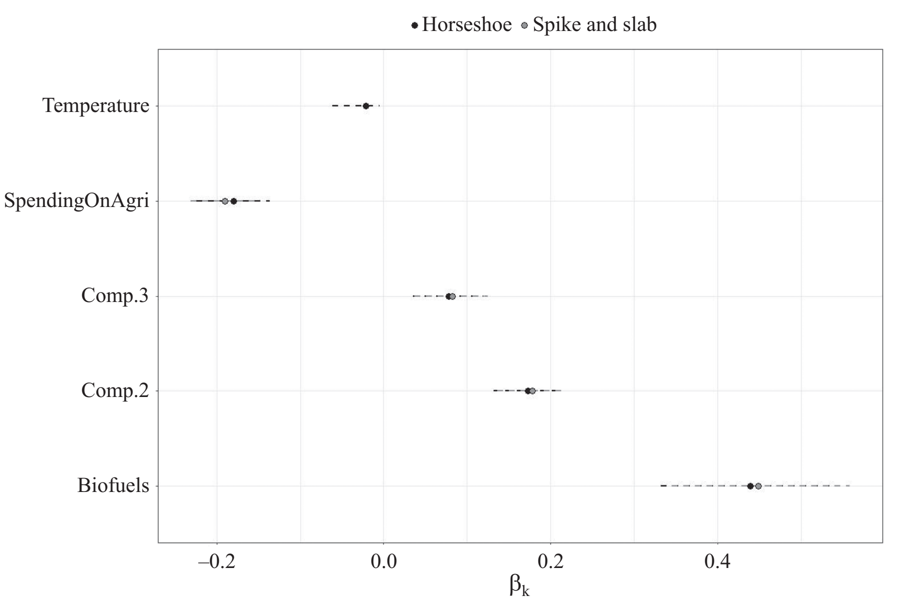


Fig. 4. Point estimates and 95% credibility intervals for β_k^* associated to the selected variables by M1 (Spike and slab) and M2 (Horseshoe).

effect and with a negative coefficient, while the coefficient associated with comp. 2, which we recall is a proxy for investment, is positive. However, the component has negative loadings on the original variables (see Figure 10b in appendix, Subsection 6.3), which means that the higher the investments (or capital intensity), the lower the losses; it also means that investing more in agriculture would reduce losses. Comp. 3 gives a small positive contribution, both for M1 and M2, which is consistent with the other energy-related variable (i.e., biofuels). Lastly, the temperature effect, selected only in M2, is the smallest. Moreover, the temperature is not measured at the crop level and is a simple yearly average. Therefore, we cannot interpret the relation as a direct effect of annual temperature on losses, but rather as the combined phenomenon by which countries with higher average temperature tend to experience smaller losses (at least in our set of data).

Recall that in Equation (5), we set a different time trend for each country. M1 estimates 24 out of 66 countries with statistically significant trends: Nine of them show an increasing trend and 15 of them show a decreasing trend. M2 instead identifies 30 countries with significant trends (22 of which are the same as in M1), with ten countries showing an increase and 20 a decrease in the food loss percentages over time. The largest estimated random effect by both M1 and M2 belongs to Malta for wheat and is equal to 1.07 on the logit scale, which corresponds to 70% percentage losses. However, the final loss factors, including the covariates' effect, are around 18%, close to the observed value in FAOSTAT. We would like to point out that Malta is an island country that imports around 90% of its wheat consumption. Loss percentages in import-dependent countries should be calculated based on domestic supply to include imports and correct extreme results. The import-dependency has been dealt with in the FLI methodology but was overlooked in this work because it was not relevant in this research context.

On the opposite side of the scale, M1 estimates the smallest effect for oats in Armenia, while M2 does so for maize (corn) in Cuba. The point estimates are -5.45 and -5.58 on the logit scale, which correspond to 0.43% and 0.37% of loss percentage, respectively (see the appendix, Subsubsection 6.5.1 for further details). In Cuba's case, the final estimates range between approximately 15% (M1) and 25% (M2), on a similar level to the country's reported losses.

The estimated values for the precision of the random effects τ^2 and the variance of the outcome are $\simeq 0.1$ and $\simeq 71$ (thus a dispersion $\phi^{-1} \simeq 0.014$), respectively, both for M1 and M2, meaning that the estimates are precise. Finally, the estimated value for the global shrinkage parameter η^2 is equal to 0.01 (for η^2, we use the Maximum A Posteriori (MAP) estimator since its posterior distribution is not symmetric. For details on variance's parameters, see the Appendix, Subsection 6.5).

4.3. Validation

To evaluate our models' predictive performance, we split the sample into training and test sets. The test set includes 953 data-points (i.e., 25% of the whole sample) and was built by removing the last five observations from the time series of each country – crop combination for time series lengths larger than eight years; only two observations were set aside for prediction purposes otherwise. We used the *Relative Mean Squared Error* (RMSE) to measure the difference between predicted and observed values. The RMSE is computed as the ratio between each model's prediction error (at the numerator) and the error that would have resulted by using the simple predictor (e.g., sample average). It can be computed as:

$$RMSE = \frac{\sum_{i=1}^{n_{te}} \left(l_i^{te} - \hat{l}_i\right)^2}{\sum_{i=1}^{n_{te}} \left(l_i^{te} - \bar{l}_{tr}\right)^2}, \tag{6}$$

where l_i^{te} are the observed losses in the test set, $\hat{l}_i$ are the predicted losses and $\bar{l}_{tr}$ is the sample average of observed losses in the training set. Predictions are obtained using the selected variables in M1 and M2, as described in Subsection. 4.2. For each model, we ran two chains with 80,000 iterations each, applied a burn-in of 40,000, and a thinning of ten, then kept 4,000 samples from each chain for inference. The overall RMSE is 0.359 and 0.358 for M1 and M2, respectively, confirming again that the two approaches are equivalent. Figure 5 shows the observed losses in the test set and their predicted values. Perfect predictions would lie on the dashed red line (or the identity line y = x). Both models show a good performance, especially for losses smaller than $\simeq$ 20% (the majority in the into data set). Besides, the average coverage of the prediction intervals for both M1 and M2 is greater than 90%. In particular, it is equal to 92.34% for M1, while it is equal to 92.55% for M2.

M1 and M2 have comparable predictive performances when the error is evaluated separately by country and commodity, although the RMSE is not uniformly distributed across countries or across commodities. In particular, Pakistan and quinoa are the country and the commodity with the highest RMSE, respectively. For Pakistan, we only have loss data for one commodity (maize), with an approximately flat time series at about 5%, as shown by the blue line in Figure 6c. The point predictions produced by our models struggle to reproduce the flat trend in this country, suggesting some unexpected behavior of one of the predictors (the spending on agriculture halved over the period) or an issue with the target variable itself; nevertheless, observed values fall into the 95% prediction

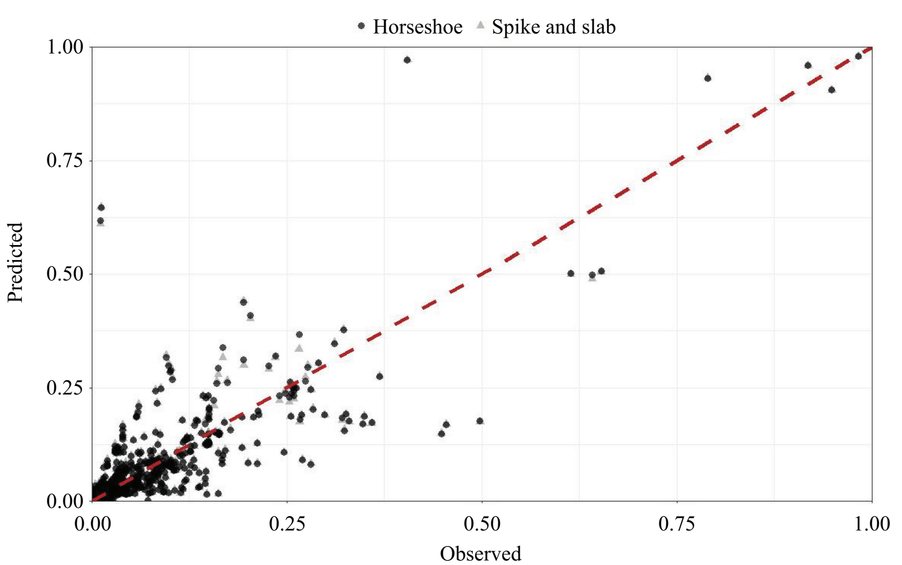
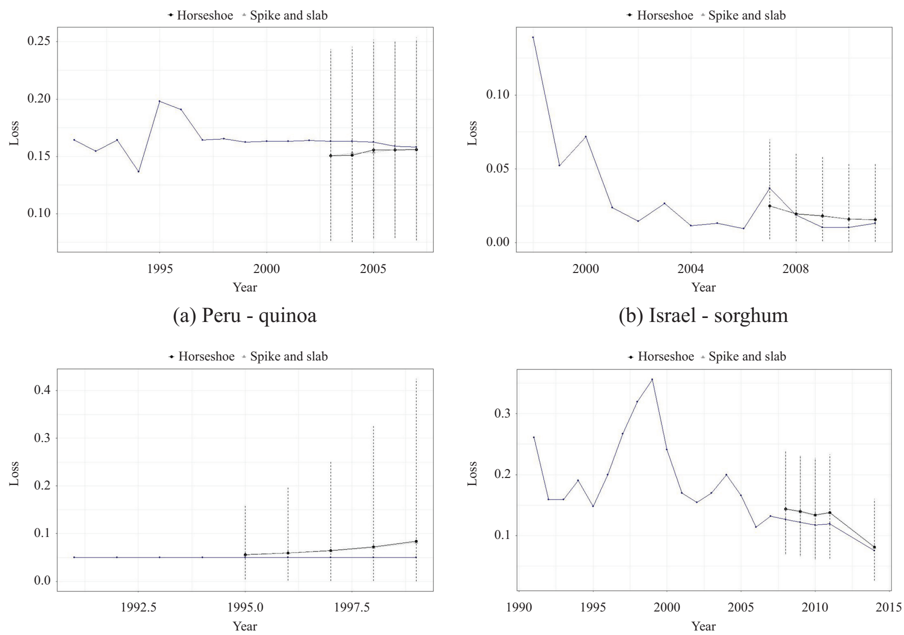


Fig. 5. Observed vs predicted losses by the two sub-models. The dashed line represents the identity line y = x.



(a) Peru - quinoa

(b) Israel - sorghum

(c) Pakistan - maize (corn)

(d) Togo - millet

Fig. 6. Observed time series and predicted values by the two sub-models with 95% predictive intervals for 4 different combinations of country – crop.

intervals for both models. Quinoa losses are only observed for Peru, as reported in Figure 6a. The models underestimate losses for this commodity; however, also, in this case, observed values fall into the 95% prediction intervals. Good prediction performances are also shown in Figure 6b and 6d: both M1 and M2 catch the *plateau* and the decreasing trend in the observed time series. For these two country-crop combinations, the prediction error is about 0.00005 for Israel-sorghum and 0.0003 for Togo-millet. Notice that the error for Israel-sorghum would have been equal to 0.0036 had we used the sorghum mean for the prediction, or 0.0058 had we predicted losses with Israel mean; for Togo-millet, the error would have been equal to 0.00037 had we used the millet mean, or equal to 0.008 had we used with Togo mean.

5. Discussion

In this work, we present a substantial improvement over previous global food losses modeling efforts. First, the proposed distributional framework is highly coherent with the nature of the data: since food losses are expressed as proportions, the Beta distribution represents a much more appropriate choice for describing their behavior. Second, the proposed approach is very flexible: the model in Equation (5) could be applied to other food groups when data will be available. Moreover, the hierarchical modeling structure covers most food loss dynamics and is easily scalable when needed (e.g., to estimate losses at supply chain step). Third, the two proposed variable selection techniques provided equivalent results. Although being computationally more demanding, the spike and slab priors allow for the computation of posterior inclusion probabilities for all the variables, providing a rigorous and straightforward way towards variables' choice. On the other hand, the horseshoe priors require the choice of *a posteriori* selection criteria, although it is considerably less demanding in terms of computational effort. Hence, it should be preferred when the latter poses a serious issue.

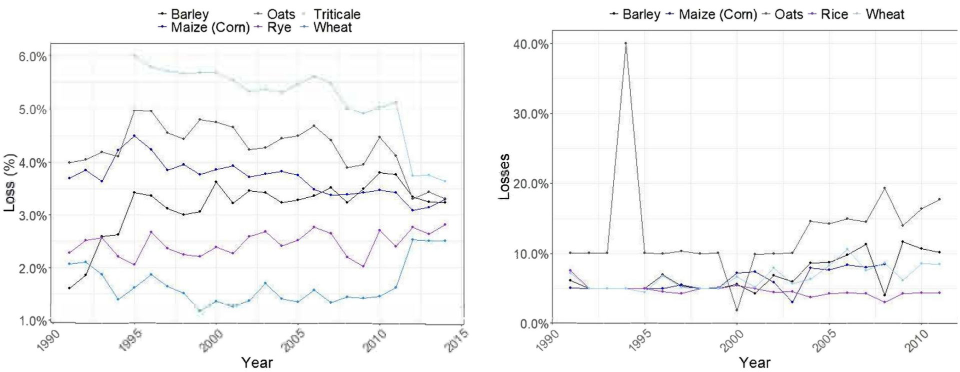
One caveat associated with the use of this model is that it is more demanding in terms of the number of observations than the hierarchical mixed-effect model developed for the SDG 12.3 methodology and the SOFA 2019 report. Our model could be developed and tested for cereals only, which account for the largest share of available data. It is not a viable option to date to compile the FLI, which needs to cover all five commodity groups, eventually with very few observations.

We would like to further remark that quality and reliability issues affect both the explaining variables and the outcome in our case study. We dealt with these issues using a Bayesian approach, which allows for the modeling of parameters' uncertainty at the prior level, but the correction of such values is not within the scope of this work and would require additional investigation. In this regard, we are aware that, in general, multiple imputation should be used, as suggested in Sinharay et al. (2001). Indeed, single imputation techniques usually underestimate the imputation process's uncertainty, and the imputed data may display a smaller variance. To handle this issue, we can consider building a model that includes a measurement error term for the imputation step. However, the increase in computational complexity does not seem to justify this solution. For these reasons, as also suggested by an anonymous reviewer, we believe that the imputation of missing data in such a context could become in itself a good method paper, hence we leave this for future developments. We also expect, when a larger amount of data will be available, to obtain the same results if we estimate the model using a maximum likelihood approach. In this work, we decided to test only Bayesian techniques because they allow to perform probabilistic uncertainty quantification in the model choice process unlike, for example, with a lasso regression. Furthermore, the lasso's optimality (theoretical properties) is only guaranteed in the framework of standard linear regression (e.g., Gaussian outcome). There is a very interesting paper by Groll et al. (2019) in which the authors propose a lasso-type penalization for generalized additive models, but in the discussion, they state that "the number of true parameters is partly overestimated." An extensive comparison between the lasso and the horseshoe is given in Bhadra et al. (2019a). Here, the authors argue that even though the lasso estimation procedure is typically computationally faster, the horseshoe prior performs better in terms of estimation thanks to its heavy tails, making it adaptive to sparse data and robust to large signals. Moreover, Polson and Scott (2010, 2012) and Datta and Ghosh (2015) have shown that horseshoe empirically outperforms lasso in terms of out-of-sample predictive sum of squares errors. Last but not least, the lack of speed can be easily overcome, as proposed in Terenin et al. (2019) and Bhadra et al. (2019b).

Overall, all the proposed models produced promising results, in terms of (1) the explanatory variables that were selected; (2) the possibility to use country-level estimates instead of clustered or global estimates; (3) the estimated trends (see FAO 2019 for comparison). More extended tests will be carried out when the data collection effort that should be undertaken by the national and international stakeholders to support policy-making towards the achievement of SDG 12.3 produces significant improvements in data availability.

6. Appendix

6.1. More Details on Data Avilability



(a) (b)

Fig. 7. Time series of available data for cereals in Austria (left panel) and Ecuador (right panel).

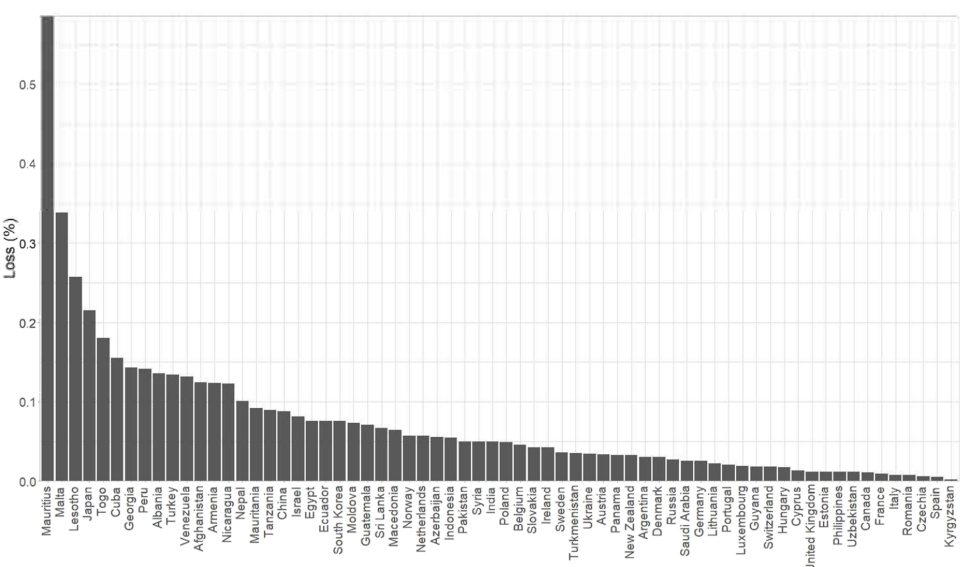


Fig. 8. Average loss percentage by country.

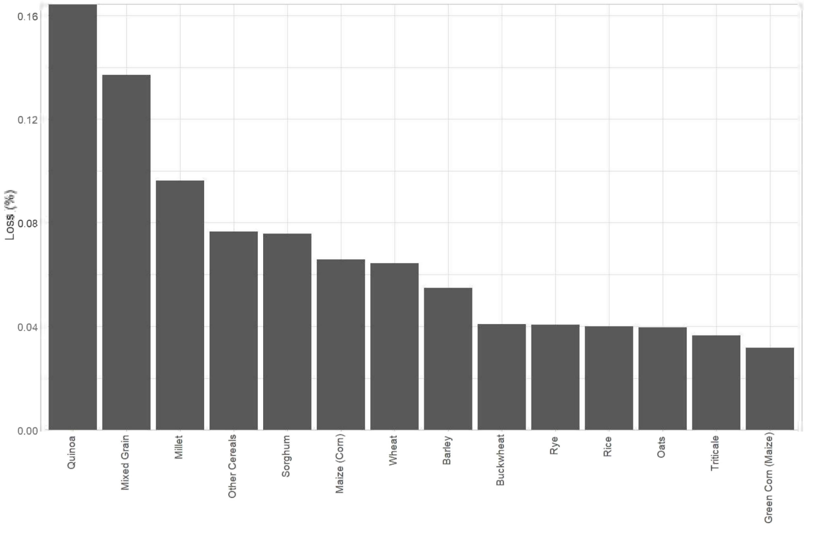


Fig. 9. Average loss percentage by crop.

6.2. Factors

Table 2. All available causal factors or proxy variables for the causes of losses, by source and category.

Variable	Source	Category	Description
Lead	World bank pink sheets	Building materials	World lead prices, annual average (nominal)
Coal	World bank pink sheets	Energy prices	World coal prices, annual average (nominal)
Copper	World bank pink sheets	Building materials	World copper prices, annual average (nominal)
Nickel	World bank pink sheets	Building materials	World nickel prices, annual average (nominal)
Crude oil	World bank pink sheets	Energy prices	World crude oil prices, annual average (nominal)
Crude petrol	World bank pink sheets	Energy prices	World crude petrol prices, annual average (nominal)
Aluminum	World bank pink sheets	Building materials	World aluminum prices, annual average (nominal)
Zinc	World bank pink sheets	Building materials	World zinc prices, annual average (nominal)
Potash	World bank pink sheets	Fertilizer	World potash prices, annual average (nominal)
Urea	World bank pink sheets	Fertilizer	World urea prices, annual average (nominal)
Phosrock	World bank pink sheets	Fertilizer	World phosrock prices, annual average (nominal)
TSP	World bank pink sheets	Fertilizer	World TSP prices, annual average (nominal)
DAP	World bank pink sheets	Fertilizer	World DAP prices, annual average (nominal)
Natural gas	International energy agency	Energy prices	World natural gas prices, annual average (nominal)

Table 2. Continued

Variable	Source	Category	Description
Gas	World bank pink sheets	Energy prices	World natural gas prices, annual average (nominal)
Heat	International energy agency	Energy prices	World natural gas prices, annual average (nominal)
Geothermal	International energy agency	Energy prices	World natural gas prices, annual average (nominal)
Oil	International energy agency	Energy prices	World natural gas prices, annual average (nominal)
Oil product	International energy agency	Energy prices	World natural gas prices, annual average (nominal)
Biofuels	International energy agency	Energy prices	World biofuels prices, annual average (nominal)
Electricity	International energy agency	Energy prices	World electricity prices, annual average (nominal)
Platinum	World bank pink sheets	Building materials	World platinum prices, annual average (nominal)
Silver	World bank pink sheets	Building materials	World silver prices, annual average (nominal)
Gold	World bank pink sheets	Building materials	World gold prices, annual average (nominal)
Iron	World bank pink sheets	Building materials	World iron prices, annual average (nominal)
Tin	World bank pink sheets	Building materials	World tni prices, annual average (nominal)
Credit to agriculture	FAOSTAT	Economic factors	Credit to agriculture
Net capital stocks	FAO agriculture capital stock database	Economic factors	Net capital stocks of agriculture, forestry and fishing
Gross fixed capital formation	FAO agriculture capital stock database	Economic factors	Gross fixed capital formation of agriculture, forestry and fishing
Gross capital stocks	FAO agriculture capital stock database	Economic factors	Gross capital stocks of agriculture, forestry and fishing
Consumption fixed capital	FAO agriculture capital ctock database	Economic factors	Consumption fixed capital of agriculture, forestry and fishing
Spending on agriculture	International food policy research insitute	Economic factors	Share of agricultural GDP
Logistic perform-ance index	World bank	Transportation and logistics	Composite indicator for evaluating trade logistics
Rainfall	World bank	Weather	Yearly average in mm
Temperature	World bank	Weather	Yearly average in degree Celsius

6.3. Factor Diagram

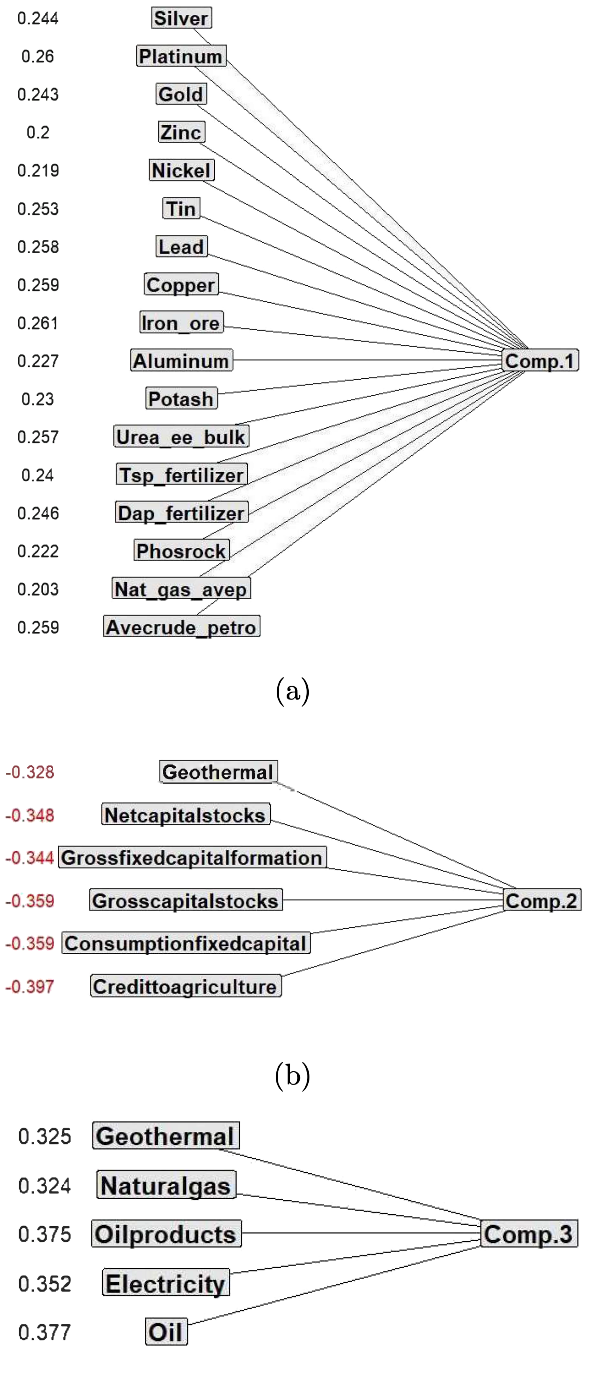


Fig. 10. Path diagram representation of PCA, showing variables associated to the 1st component (a), to the 2nd component (b) and to the third component (c).

6.4. Example Codes

6.4.1. Spike and Slab Implementation

Listing 1. Spike and slab estimation with JAGS

```
model {
# Likelihood
for(i in 1:n){
 Y[i] ~ dbeta(alpha[i], delta[i])
 alpha[i] <- mu[i] * phi
 delta[i]  <- (1-mu[i]) * phi
 logLik[i] <- (alpha[i]-1)*log(Y[i]) + (delta[i]-1)*log(1-Y[i])
 logit(mu[i]) <- zbetaYear[CTRY[i]]*Year[i] +
                    inprod(X[i,], zbeta[]) +
                    etacoucomm[CC[i]]
 }

# Priors country-commodity
for(k in 1:neff){
 etacoucomm[k]~dnorm(0,tau)
 }

# Prior for tau
 tau ~ dgamma(4, 0.1)
# Prior for phi
 phi ~ dunif(5,150)

# Prior for beta and gamma
for(ctry in 1:nctry){
 ppYear[ctry] ~ dbeta(5,5)
 gammaYear[ctry] ~  dbern(ppYear[ctry])
 betaYear[ctry] ~ dnorm(0,0.001)
 zbetaYear[ctry] = betaYear[ctry]*gammaYear[ctry]
}

for(j in 1:p){
 beta[j] ~ dnorm(0,0.001)
 pp[j] ~ dbeta(5,5)
 gamma[j] ~  dbern(pp[j])
 zbeta[j] =  beta[j]*gamma[j]

 }
}
```

Listing 2. Spike and slab prediction with JAGS

```
model {
# Train
for(i in 1:ntrain){
 Ytrain[i]  ~ dbeta(alphatrain[i], deltatrain[i])
 alphatrain[i] <- mutrain[i] * phi
 deltatrain[i]  <- (1-mutrain[i]) * phi
 logit(mutrain[i]) <- betaYear[CTRYtrain[i]]*Yeartrain[i] +
                      inprod(Xtrain[i,],beta[]) +
                      etacoucomm[CCtrain[i]]
 }

# Prediction - test
for(i in 1:ntest){
 Ytest[i]  ~ dbeta(alphatest[i], deltatest[i])
 alphatest[i] <- mutest[i] * phi
 deltatest[i]  <- (1-mutest[i]) * phi
 logit(mutest[i]) <- betaYear[CTRYtest[i]]*Yeartest[i] +
                     inprod(Xtest[i,],beta[]) +
                     etacoucomm[CCtest[i]]
 }

for(k in 1:neff){
        etacoucomm[k]~dnorm(0,tau)
 }

# Prior for tau
 tau ~ dgamma(4, 0.1)
# Prior for phi
 phi ~ dunif(5,150)

# Prior for beta and gamma
for(ctry in 1:nctry){
  betaYear[ctry] ~ dnorm(0,0.001)
 }
for(j in 1:p){
  beta[j] ~ dnorm(0,0.001)
 }
}
```

6.4.2. Horseshoe Implementation

Listing 3. Horseshoe estimation with JAGS

```
model {
# Likelihood
for(i in 1:n){
 Y[i] ~ dbeta(alpha[i], delta[i])
 alpha[i] <- mu[i] * phi
 delta[i]  <- (1-mu[i]) * phi
 logLik[i] <- (alpha[i]-1)*log(Y[i]) + (delta[i]-1)*log(1-Y[i])
 logit(mu[i]) <- betaYear[CTRY[i]]*Year[i] +
                    inprod(X[i,], beta[]) +
                    etacoucomm[CC[i]]
 }
# Priors country-commodity
for(k in 1:neff){
        etacoucomm[k]~dnorm(0,tau)
 }

# Prior for beta and gamma
for(ctry in 1:nctry){
  shrinkYear[ctry] ~ dt(0,1,1)T(0,)
  betaYear[ctry] ~ dnorm(0, 1/(shrinkYear[ctry]*global))
 }
for(j in 1:p){
        shrink[j] ~ dt(0,1,1)T(0,)
        beta[j] ~ dnorm(0, 1/(shrink[j]*global))
 }

# Prior for tau
 tau ~ dgamma(4, 0.1)
# Prior for phi
 phi ~ dunif(5,150)
# Prior for global shrinkage
        global ~ dt(0,1,1)T(0,)
}
```

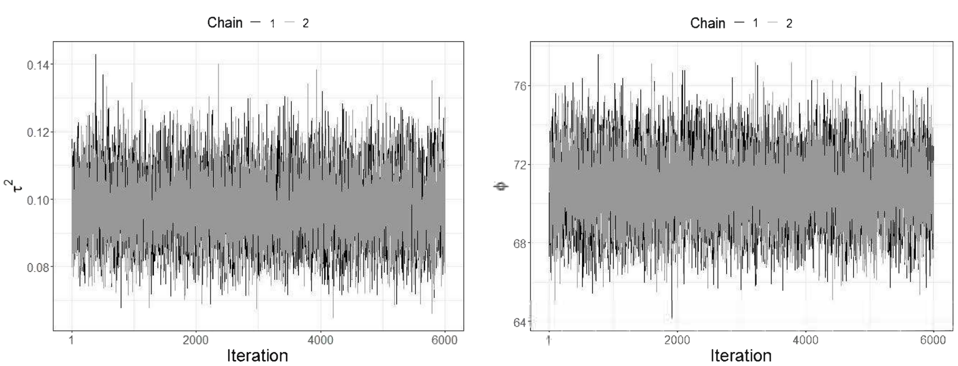
Listing 4. Horseshoe prediction with JAGS

```
model {
# Train
for(i in 1:ntrain){
 Ytrain[i]  ~ dbeta(alphatrain[i], deltatrain[i])
 alphatrain[i] <- mutrain[i] * phi
 deltatrain[i]  <- (1-mutrain[i]) * phi
 logit(mutrain[i]) <- betaYear[CTRYtrain[i]]*Yeartrain[i] +
                      inprod(Xtrain[i,],beta[]) +
                      etacoucomm[CCtrain[i]]
 }

# Prediction - test
for(i in 1:ntest){
 Ytest[i]  ~ dbeta(alphatest[i], deltatest[i])
 alphatest[i] <- mutest[i] * phi
 deltatest[i]  <- (1-mutest[i]) * phi
 logit(mutest[i]) <- betaYear[CTRYtest[i]]*Yeartest[i] +
                     inprod(Xtest[i,],beta[]) +
                     etacoucomm[CCtest[i]]
 }
## priors country-commodity
for(k in 1:neff){
  etacoucomm[k]~dnorm(0,tau)
 }

# Prior for beta and gamma
for(ctry in 1:nctry){
 betaYear[ctry] ~ dnorm(0, 0.001)
 }
for(j in 1:p){
  shrink[j] ~ dt(0,1,1)T(0,)
  beta[j] ~ dnorm(0, 1/(shrink[j]*global))
 }

# Prior for tau
 tau ~ dgamma(4, 0.1)
# Prior for phi
 phi ~ dunif(5,150)
# Prior for global shrinkage
 global ~ dt(0,1,1)T(0,)
}
```

6.5. *Estimation And Diagnostics*

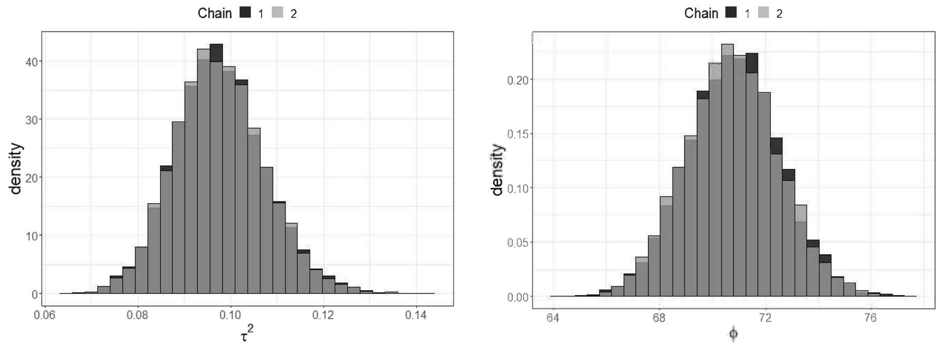
Table 3. Selected variables with the two different priors and 95% posterior credibility intervals of the β_k^**.*

Model	Variable	$\hat{\gamma}$	$q_{.025}$	Mean	$q_{.975}$
M1	*Biofuels*	1	0.345	0.449	0.558
	SpendingOnAgri	1	−0.231	−0.190	−0.149
	Comp.2	1	0.136	0.178	0.219
	Comp.3	0.56	0.037	0.081	0.125
M2	*Biofuels*		0.332	0.439	0.547
	SpendingOnAgri		−0.224	−0.180	−0.136
	Comp.2		0.132	0.173	0.213
	Comp.3		0.035	0.078	0.120
	Temperature		−0.0617	−0.021	−0.005



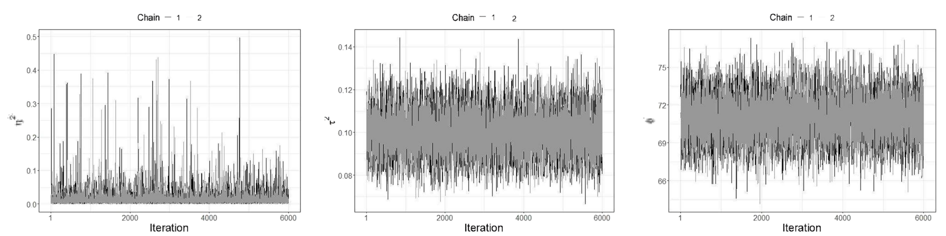
(a) (b)

Fig. 11. Traceplots of the estimated variance components by M1: variance of the random effects τ^2 *(a) and variance of the outcome* ϕ *(b).*



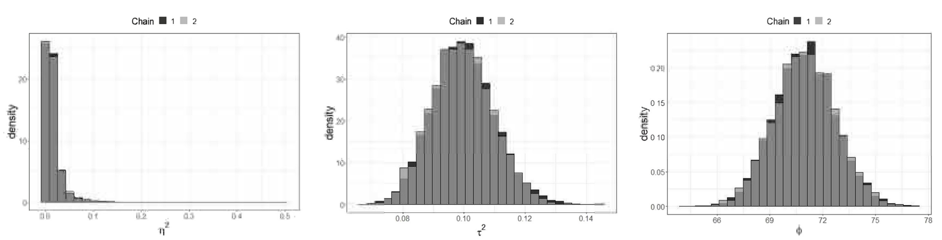
(a) (b)

Fig. 12. Posterior density of the estimated variance components by M1: variance of the random effects τ^2 (a) and variance of the outcome ϕ (b).



(a) (b) (c)

Fig. 13. Traceplots of the estimated variance components by M2: global shrinkage parameter η^2 (a), variance of the random effects τ^2 (b) and variance of the outcome ϕ(c).



(a) (b) (c)

Fig. 14. Posterior density of the estimated variance components by M2: global shrinkage parameter η^2 (a), variance of the random effects τ^2 (b) and variance of the outcome ϕ (c).

6.5.1. Country-Crop Random Effects

Table 4. Point estimates and 95% credible intervals for the country-crop random effects by M1.

Country	Crop	$q_{.025}$	Mean	$q_{.975}$
Afghanistan	wheat	−2.77	−2.44	−2.14
Albania	wheat	−2.82	−2.50	−2.20
Argentina	barley	−3.73	−3.32	−2.96
Argentina	maize (corn)	−4.39	−3.86	−3.39
Argentina	oats	−3.51	−3.13	−2.80
Argentina	rice	−4.27	−3.80	−3.38
Argentina	rye	−4.18	−3.71	−3.29
Armenia	barley	−5.05	−4.33	−3.65
Armenia	maize (corn)	−3.29	−2.90	−2.56
Armenia	oats	−6.54	−5.46	−4.52
Armenia	other cereals	−4.68	−4.12	−3.60
Armenia	wheat	−5.48	−4.78	−4.11
Austria	barley	−4.51	−4.10	−3.70
Austria	maize (corn)	−4.28	−3.88	−3.49
Austria	oats	−4.14	−3.75	−3.37
Austria	rye	−4.71	−4.29	−3.87
Austria	triticale	−3.93	−3.51	−3.11
Austria	wheat	−5.08	−4.63	−4.20
Azerbaijan	barley	−4.34	−3.94	−3.57
Azerbaijan	maize (corn)	−3.97	−3.61	−3.27
Azerbaijan	rice	−1.80	−1.55	−1.31
Azerbaijan	wheat	−4.33	−3.93	−3.56
Belgium	barley	−5.24	−4.35	−3.49
Belgium	maize (corn)	−5.59	−4.67	−3.76
Belgium	oats	−5.54	−4.61	−3.73
Belgium	rye	−3.13	−2.38	−1.64
Belgium	wheat	−5.47	−4.53	−3.65
Canada	barley	−6.15	−5.45	−4.84
Canada	green corn (maize)	−3.64	−3.12	−2.66
Canada	oats	−5.80	−5.10	−4.48
Canada	rye	−4.41	−3.80	−3.28
Canada	wheat	−6.11	−5.38	−4.70
China	maize (corn)	−3.05	−2.43	−1.89
Cuba	maize (corn)	−8.72	−5.42	−2.63
Cyprus	barley	−5.55	−4.72	−4.00
Cyprus	wheat	−4.42	−3.61	−2.88
Czechia	barley	−5.03	−4.49	−4.00
Czechia	oats	−5.28	−4.72	−4.23
Czechia	rye	−5.11	−4.54	−4.02
Czechia	wheat	−5.30	−4.76	−4.27
Denmark	barley	−3.31	−2.94	−2.58
Denmark	oats	−3.31	−2.92	−2.55
Denmark	rye	−3.31	−2.93	−2.59
Denmark	wheat	−3.43	−3.04	−2.68
Ecuador	barley	−2.78	−2.50	−2.23
Ecuador	maize (corn)	−2.93	−2.63	−2.33
Ecuador	oats	−2.16	−1.90	−1.66

Table 4. Continued

Country	Crop	$q_{.025}$	Mean	$q_{.975}$
Ecuador	rice	− 3.03	− 2.72	− 2.43
Ecuador	wheat	− 2.77	− 2.49	− 2.23
Egypt	barley	− 3.18	− 2.81	− 2.47
Estonia	barley	− 4.76	− 4.20	− 3.68
Estonia	rye	− 4.39	− 3.89	− 3.43
France	barley	− 3.95	− 3.47	− 3.03
France	maize (corn)	− 3.84	− 3.38	− 2.93
France	oats	− 3.91	− 3.42	− 2.96
France	rice	− 3.66	− 3.21	− 2.76
France	wheat	− 3.94	− 3.47	− 3.03
Georgia	maize (corn)	− 3.30	− 2.99	− 2.70
Georgia	wheat	− 1.61	− 1.37	− 1.14
Germany	barley	− 2.62	− 2.14	− 1.67
Germany	maize (corn)	− 2.43	− 1.95	− 1.49
Germany	oats	− 2.59	− 2.11	− 1.64
Germany	rye	− 2.53	− 2.05	− 1.60
Germany	triticale	− 2.80	− 2.31	− 1.83
Germany	wheat	− 2.64	− 2.15	− 1.68
Guatemala	maize (corn)	− 4.49	− 4.11	− 3.76
Guatemala	rice	− 3.94	− 3.61	− 3.31
Guatemala	sorghum	− 4.71	− 4.32	− 3.97
Guatemala	wheat	− 3.33	− 3.09	− 2.84
Guyana	rice	− 5.25	− 3.97	− 3.05
Hungary	maize (corn)	− 4.27	− 3.87	− 3.49
Hungary	oats	− 4.43	− 4.00	− 3.61
Hungary	rice	− 3.70	− 3.19	− 2.70
Hungary	rye	− 3.89	− 3.50	− 3.13
Hungary	wheat	− 4.00	− 3.59	− 3.22
India	barley	− 3.64	− 2.80	− 2.09
India	maize (corn)	− 1.81	− 1.26	− 0.75
India	millet	− 2.85	− 2.17	− 1.58
India	sorghum	− 2.42	− 1.81	− 1.25
India	wheat	− 3.20	− 2.45	− 1.80
Indonesia	maize (corn)	− 3.00	− 2.66	− 2.35
Indonesia	rice	− 3.01	− 2.67	− 2.35
Ireland	barley	− 3.04	− 2.72	− 2.41
Ireland	oats	− 3.17	− 2.84	− 2.54
Ireland	wheat	− 2.81	− 2.50	− 2.20
Israel	barley	− 1.28	− 1.04	− 0.80
Israel	maize (corn)	− 3.59	− 3.25	− 2.95
Israel	sorghum	− 3.08	− 2.56	− 2.10
Italy	barley	− 4.56	− 3.93	− 3.32
Italy	maize (corn)	− 4.73	− 4.08	− 3.47
Italy	oats	− 3.28	− 2.74	− 2.22
Italy	sorghum	− 3.70	− 2.88	− 2.21
Italy	wheat	− 3.83	− 3.25	− 2.70
Japan	barley	− 1.28	− 0.68	− 0.07
Japan	rice	− 4.86	− 4.15	− 3.45
Japan	wheat	− 2.17	− 1.53	− 0.87

Table 4. Continued

Country	Crop	$q_{.025}$	Mean	$q_{.975}$
Kyrgyzstan	rice	−6.05	−5.12	−4.32
Lesotho	maize (corn)	−1.61	−1.17	−0.76
Lithuania	barley	−4.47	−4.02	−3.60
Lithuania	wheat	−3.96	−3.44	−2.96
Luxembourg	wheat	−6.13	−3.13	−0.32
Macedonia	maize (corn)	−2.68	−2.27	−1.89
Macedonia	rye	−2.51	−2.16	−1.83
Malta	wheat	0.82	1.07	1.34
Mauritania	sorghum	−2.69	−2.31	−1.95
Mauritius	maize (corn)	−1.32	−0.97	−0.65
Moldova	wheat	−2.52	−1.98	−1.49
Nepal	barley	−2.31	−2.07	−1.83
Nepal	maize (corn)	−2.35	−2.10	−1.87
Nepal	wheat	−2.32	−2.07	−1.83
Netherlands	barley	−3.40	−3.06	−2.73
Netherlands	maize (corn)	−2.40	−2.10	−1.81
Netherlands	rye	−3.23	−2.77	−2.36
Netherlands	triticale	−3.07	−2.56	−2.12
Netherlands	wheat	−4.01	−3.64	−3.28
New Zealand	barley	−3.90	−3.45	−3.04
New Zealand	oats	−3.58	−3.17	−2.80
Nicaragua	maize (corn)	−1.48	−1.23	−0.99
Nicaragua	rice	−3.76	−3.38	−3.04
Norway	barley	−3.23	−2.92	−2.64
Norway	oats	−3.67	−3.33	−3.03
Norway	rye	−2.94	−2.65	−2.38
Norway	wheat	−2.42	−2.17	−1.92
Pakistan	maize (corn)	−3.84	−2.93	−2.22
Panama	maize (corn)	−3.10	−2.75	−2.42
Panama	rice	−3.25	−2.89	−2.55
Peru	barley	−1.19	−1.00	−0.80
Peru	maize (corn)	−3.15	−2.84	−2.56
Peru	quinoa	−1.83	−1.61	−1.40
Peru	rice	−4.15	−3.58	−3.08
Peru	rye	−3.12	−2.70	−2.34
Peru	wheat	−1.55	−1.34	−1.14
Philippines	maize (corn)	−4.59	−4.07	−3.59
Philippines	rice	−4.29	−3.78	−3.32
Poland	barley	−3.04	−2.76	−2.49
Poland	buckwheat	−3.27	−2.97	−2.69
Poland	maize (corn)	−3.01	−2.73	−2.46
Poland	oats	−2.75	−2.49	−2.24
Poland	rye	−3.23	−2.94	−2.65
Poland	triticale	−3.21	−2.91	−2.64
Poland	wheat	−3.05	−2.76	−2.50
Portugal	maize (corn)	−4.55	−4.07	−3.63
Portugal	rice	−5.93	−5.36	−4.84
Portugal	rye	−4.18	−3.76	−3.38
Portugal	wheat	−4.93	−4.48	−4.06

Table 4. Continued

Country	Crop	$q_{.025}$	Mean	$q_{.975}$
Romania	maize (corn)	− 4.79	− 4.23	− 3.73
Romania	wheat	− 5.00	− 4.44	− 3.93
Russia	barley	− 3.61	− 3.11	− 2.62
Russia	maize (corn)	− 3.47	− 2.95	− 2.44
Russia	other cereals	− 2.03	− 1.59	− 1.16
Russia	rice	− 3.40	− 2.92	− 2.44
Russia	wheat	− 3.98	− 3.45	− 2.93
Saudi Arabia	wheat	− 4.42	− 3.80	− 3.25
Slovakia	barley	− 3.29	− 2.95	− 2.64
Slovakia	maize (corn)	− 2.66	− 2.39	− 2.13
Slovakia	oats	− 3.60	− 3.26	− 2.95
Slovakia	rye	− 3.66	− 3.29	− 2.95
Slovakia	wheat	− 3.77	− 3.40	− 3.06
South Korea	rice	− 3.71	− 3.26	− 2.86
Spain	barley	− 5.11	− 4.20	− 3.43
Spain	maize (corn)	− 4.72	− 4.09	− 3.48
Spain	rye	− 4.59	− 4.01	− 3.47
Spain	wheat	− 4.71	− 4.08	− 3.47
Sri Lanka	maize (corn)	− 2.50	− 2.25	− 2.02
Sri Lanka	millet	− 3.61	− 3.31	− 3.02
Sri Lanka	rice	− 2.98	− 2.70	− 2.43
Sweden	barley	− 3.45	− 3.11	− 2.79
Sweden	mixed grain	− 3.08	− 2.76	− 2.48
Sweden	rye	− 3.02	− 2.70	− 2.39
Sweden	wheat	− 3.41	− 3.07	− 2.76
Switzerland	barley	− 3.67	− 3.28	− 2.92
Switzerland	maize (corn)	− 3.70	− 3.31	− 2.94
Switzerland	oats	− 3.66	− 3.27	− 2.91
Switzerland	rye	− 3.57	− 3.19	− 2.85
Switzerland	triticale	− 4.21	− 3.44	− 2.77
Switzerland	wheat	− 3.59	− 3.21	− 2.87
Syria	wheat	− 3.35	− 2.88	− 2.46
Tanzania	maize (corn)	− 3.43	− 2.99	− 2.60
Toga	millet	− 1.58	− 1.34	− 1.10
Turkey	mixed grain	− 1.63	− 1.45	− 1.27
Turkey	oats	− 2.01	− 1.81	− 1.62
Turkey	rice	− 3.48	− 3.18	− 2.91
Turkey	rye	− 2.01	− 1.81	− 1.62
Turkey	wheat	− 2.54	− 2.32	− 2.10
Turkmenistan	rice	− 4.23	− 3.42	− 2.77
Ukraine	maize (corn)	− 3.21	− 2.88	− 2.57
Ukraine	oats	− 3.21	− 2.90	− 2.61
Ukraine	rice	− 3.93	− 3.56	− 3.21
Ukraine	rye	− 3.59	− 3.22	− 2.88
United Kingdom	barley	− 4.73	− 4.22	− 3.76
United Kingdom	oats	− 4.74	− 4.22	− 3.74
United Kingdom	wheat	− 4.10	− 3.67	− 3.27
Uzbekistan	maize (corn)	− 4.99	− 4.01	− 3.21
Venezuela	maize (corn)	− 1.66	− 1.46	− 1.25
Venezuela	rice	− 2.35	− 2.09	− 1.84
Venezuela	sorghum	− 1.55	− 1.35	− 1.16

Table 5. Point estimates and 95% credible intervals for the country-crop random effects by M2.

Country	Crop	$q_{.025}$	Mean	$q_{.975}$
Afghanistan	wheat	−2.76	−2.45	−2.15
Albania	wheat	−2.84	−2.51	−2.21
Argentina	barley	−3.74	−3.33	−2.97
Argentina	maize (corn)	−4.40	−3.86	−3.39
Argentina	oats	−3.51	−3.14	−2.81
Argentina	rice	−4.29	−3.81	−3.38
Argentina	rye	−4.18	−3.71	−3.30
Armenia	barley	−5.02	−4.31	−3.63
Armenia	maize (corn)	−3.25	−2.88	−2.55
Armenia	oats	−6.55	−5.43	−4.50
Armenia	other cereals	−4.65	−4.10	−3.57
Armenia	wheat	−5.46	−4.76	−4.09
Austria	barley	−4.47	−4.05	−3.65
Austria	maize (corn)	−4.24	−3.84	−3.45
Austria	oats	−4.10	−3.71	−3.32
Austria	rye	−4.67	−4.24	−3.84
Austria	triticale	−3.87	−3.46	−3.06
Austria	wheat	−5.04	−4.59	−4.14
Azerbaijan	barley	−4.34	−3.93	−3.56
Azerbaijan	maize (corn)	−3.97	−3.61	−3.27
Azerbaijan	rice	−1.80	−1.55	−1.31
Azerbaijan	wheat	−4.33	−3.93	−3.56
Belgium	barley	−5.19	−4.31	−3.47
Belgium	maize (corn)	−5.56	−4.63	−3.74
Belgium	oats	−5.47	−4.57	−3.68
Belgium	rye	−3.10	−2.34	−1.61
Belgium	wheat	−5.44	−4.50	−3.61
Canada	barley	−6.10	−5.42	−4.81
Canada	green corn (maize)	−3.61	−3.09	−2.63
Canada	oats	−5.78	−5.07	−4.49
Canada	rye	−4.38	−3.78	−3.26
Canada	wheat	−6.09	−5.36	−4.69
China	maize (corn)	−3.05	−2.44	−1.91
Cuba	maize (corn)	−8.44	−5.58	−2.71
Cyprus	barley	−5.53	−4.72	−4.00
Cyprus	wheat	−4.42	−3.61	−2.86
Czechia	barley	−5.03	−4.51	−4.02
Czechia	oats	−5.28	−4.74	−4.26
Czechia	rye	−5.12	−4.56	−4.04
Czechia	wheat	−5.32	−4.77	−4.28
Denmark	barley	−3.32	−2.92	−2.55
Denmark	oats	−3.30	−2.90	−2.52
Denmark	rye	−3.31	−2.92	−2.55
Denmark	wheat	−3.41	−3.02	−2.67
Ecuador	barley	−2.80	−2.51	−2.24
Ecuador	maize (corn)	−2.95	−2.64	−2.35
Ecuador	oats	−2.16	−1.91	−1.67
Ecuador	rice	−3.02	−2.72	−2.43
Ecuador	wheat	−2.78	−2.50	−2.23
Egypt	barley	−3.19	−2.82	−2.48

Table 5. Continued

Country	Crop	$q_{.025}$	Mean	$q_{.975}$
Estonia	barley	−4.79	−4.22	−3.71
Estonia	rye	−4.40	−3.90	−3.44
France	barley	−3.98	−3.51	−3.07
France	maize (corn)	−3.90	−3.42	−2.99
France	oats	−3.96	−3.46	−3.00
France	rice	−3.70	−3.24	−2.81
France	wheat	−3.97	−3.51	−3.07
Georgia	maize (corn)	−3.32	−3.01	−2.72
Georgia	wheat	−1.64	−1.39	−1.16
Germany	barley	−2.69	−2.20	−1.73
Germany	maize (corn)	−2.49	−2.01	−1.54
Germany	oats	−2.64	−2.17	−1.68
Germany	rye	−2.60	−2.11	−1.64
Germany	triticale	−2.89	−2.37	−1.89
Germany	wheat	−2.71	−2.21	−1.73
Guatemala	maize (corn)	−4.49	−4.11	−3.76
Guatemala	rice	−3.93	−3.60	−3.29
Guatemala	sorghum	−4.71	−4.32	−3.96
Guatemala	wheat	−3.33	−3.09	−2.85
Guyana	rice	−5.25	−3.98	−3.05
Hungary	maize (corn)	−4.30	−3.89	−3.50
Hungary	oats	−4.47	−4.02	−3.62
Hungary	rice	−3.73	−3.21	−2.73
Hungary	rye	−3.89	−3.51	−3.14
Hungary	wheat	−4.02	−3.61	−3.23
India	barley	−3.63	−2.80	−2.11
India	maize (corn)	−1.82	−1.26	−0.74
India	millet	−2.85	−2.17	−1.56
India	sorghum	−2.43	−1.82	−1.25
India	wheat	−3.22	−2.45	−1.81
Indonesia	maize (corn)	−2.95	−2.61	−2.28
Indonesia	rice	−2.97	−2.62	−2.28
Ireland	barley	−3.00	−2.67	−2.34
Ireland	oats	−3.14	−2.80	−2.47
Ireland	wheat	−2.78	−2.45	−2.14
Israel	barley	−1.28	−1.04	−0.81
Israel	maize (corn)	−3.60	−3.26	−2.95
Israel	sorghum	−3.08	−2.56	−2.10
Italy	barley	−4.63	−3.99	−3.37
Italy	maize (corn)	−4.79	−4.14	−3.53
Italy	oats	−3.33	−2.80	−2.29
Italy	sorghum	−3.75	−2.95	−2.28
Italy	wheat	−3.89	−3.31	−2.75
Japan	barley	−1.36	−0.77	−0.17
Japan	rice	−4.94	−4.24	−3.55
Japan	wheat	−2.25	−1.62	−0.97
Kyrgyzstan	rice	−6.03	−5.11	−4.30
Lesotho	maize (corn)	−1.62	−1.17	−0.76
Lithuania	barley	−4.53	−4.07	−3.64

Table 5. Continued

Country	Crop	$q_{.025}$	Mean	$q_{.975}$
Lithuania	wheat	−4.04	−3.49	−3.00
Luxembourg	wheat	−6.05	−3.10	−0.14
Macedonia	maize (corn)	−2.68	−2.29	−1.90
Macedonia	rye	−2.51	−2.17	−1.85
Malta	wheat	0.81	1.06	1.32
Mauritania	sorghum	−2.68	−2.31	−1.97
Mauritius	maize (corn)	−1.29	−0.95	−0.63
Moldova	wheat	−2.53	−1.97	−1.46
Nepal	barley	−2.35	−2.10	−1.86
Nepal	maize (corn)	−2.38	−2.14	−1.90
Nepal	wheat	−2.35	−2.10	−1.86
Netherlands	barley	−3.41	−3.07	−2.74
Netherlands	maize (corn)	−2.41	−2.11	−1.82
Netherlands	rye	−3.24	−2.78	−2.37
Netherlands	triticale	−3.08	−2.57	−2.13
Netherlands	wheat	−4.03	−3.65	−3.28
New Zealand	barley	−3.92	−3.46	−3.05
New Zealand	oats	−3.59	−3.18	−2.80
Nicaragua	maize (corn)	−1.45	−1.19	−0.94
Nicaragua	rice	−3.72	−3.34	−2.98
Norway	barley	−3.22	−2.92	−2.63
Norway	oats	−3.66	−3.33	−3.02
Norway	rye	−2.94	−2.64	−2.37
Norway	wheat	−2.42	−2.16	−1.92
Pakistan	maize (corn)	−3.87	−2.93	−2.22
Panama	maize (corn)	−3.08	−2.72	−2.39
Panama	rice	−3.23	−2.86	−2.52
Peru	barley	−1.21	−1.01	−0.81
Peru	maize (corn)	−3.15	−2.85	−2.57
Peru	quinoa	−1.84	−1.62	−1.40
Peru	rice	−4.17	−3.59	−3.09
Peru	rye	−3.13	−2.71	−2.35
Peru	wheat	−1.56	−1.35	−1.15
Philippines	maize (corn)	−4.56	−4.02	−3.55
Philippines	rice	−4.26	−3.74	−3.28
Poland	barley	−3.03	−2.75	−2.49
Poland	buckwheat	−3.26	−2.96	−2.68
Poland	maize (corn)	−3.01	−2.73	−2.46
Poland	oats	−2.75	−2.49	−2.24
Poland	rye	−3.23	−2.94	−2.66
Poland	triticale	−3.20	−2.91	−2.63
Poland	wheat	−3.03	−2.76	−2.50
Portugal	maize (corn)	−4.53	−4.06	−3.62
Portugal	rice	−5.95	−5.37	−4.84
Portugal	rye	−4.19	−3.76	−3.37
Portugal	wheat	−4.93	−4.47	−4.05
Romania	maize (corn)	−4.79	−4.23	−3.74
Romania	wheat	−5.03	−4.46	−3.95
Russia	barley	−3.65	−3.13	−2.65

Table 5. Continued

Country	Crop	$q_{.025}$	Mean	$q_{.975}$
Russia	maize (corn)	−3.48	−2.98	−2.48
Russia	other cereals	−2.04	−1.61	−1.19
Russia	rice	−3.43	−2.94	−2.47
Russia	wheat	−4.00	−3.46	−2.96
Saudi Arabia	wheat	−4.45	−3.81	−3.27
Slovakia	barley	−3.31	−2.97	−2.64
Slovakia	maize (corn)	−2.67	−2.40	−2.16
Slovakia	oats	−3.61	−3.27	−2.96
Slovakia	rye	−3.67	−3.30	−2.96
Slovakia	wheat	−3.78	−3.41	−3.08
South Korea	rice	−3.72	−3.27	−2.87
Spain	barley	−5.12	−4.22	−3.46
Spain	maize (corn)	−4.74	−4.11	−3.51
Spain	rye	−4.60	−4.03	−3.49
Spain	wheat	−4.72	−4.10	−3.52
Sri Lanka	maize (corn)	−2.46	−2.20	−1.94
Sri Lanka	millet	−3.58	−3.25	−2.93
Sri Lanka	rice	−2.94	−2.64	−2.36
Sweden	barley	−3.47	−3.12	−2.79
Sweden	mixed grain	−3.08	−2.77	−2.48
Sweden	rye	−3.03	−2.71	−2.41
Sweden	wheat	−3.40	−3.08	−2.77
Switzerland	barley	−3.67	−3.29	−2.92
Switzerland	maize (corn)	−3.71	−3.31	−2.95
Switzerland	oats	−3.65	−3.27	−2.91
Switzerland	rye	−3.57	−3.20	−2.87
Switzerland	triticale	−4.21	−3.45	−2.79
Switzerland	wheat	−3.59	−3.22	−2.88
Syria	wheat	−3.33	−2.85	−2.44
Tanzania	maize (corn)	−3.44	−3.01	−2.62
Toga	millet	−1.59	−1.34	−1.10
Turkey	mixed grain	−1.64	−1.46	−1.28
Turkey	oats	−2.01	−1.82	−1.62
Turkey	rice	−3.48	−3.19	−2.92
Turkey	rye	−2.02	−1.82	−1.63
Turkey	wheat	−2.55	−2.32	−2.11
Turkmenistan	rice	−4.22	−3.43	−2.77
Ukraine	maize (corn)	−3.20	−2.87	−2.57
Ukraine	oats	−3.21	−2.89	−2.59
Ukraine	rice	−3.94	−3.55	−3.20
Ukraine	rye	−3.58	−3.21	−2.86
United Kingdom	barley	−4.72	−4.22	−3.76
United Kingdom	oats	−4.74	−4.22	−3.75
United Kingdom	wheat	−4.10	−3.67	−3.27
Uzbekistan	maize (corn)	−5.00	−4.02	−3.22
Venezuela	maize (corn)	−1.68	−1.47	−1.26
Venezuela	rice	−2.37	−2.10	−1.84
Venezuela	sorghum	−1.57	−1.36	−1.16

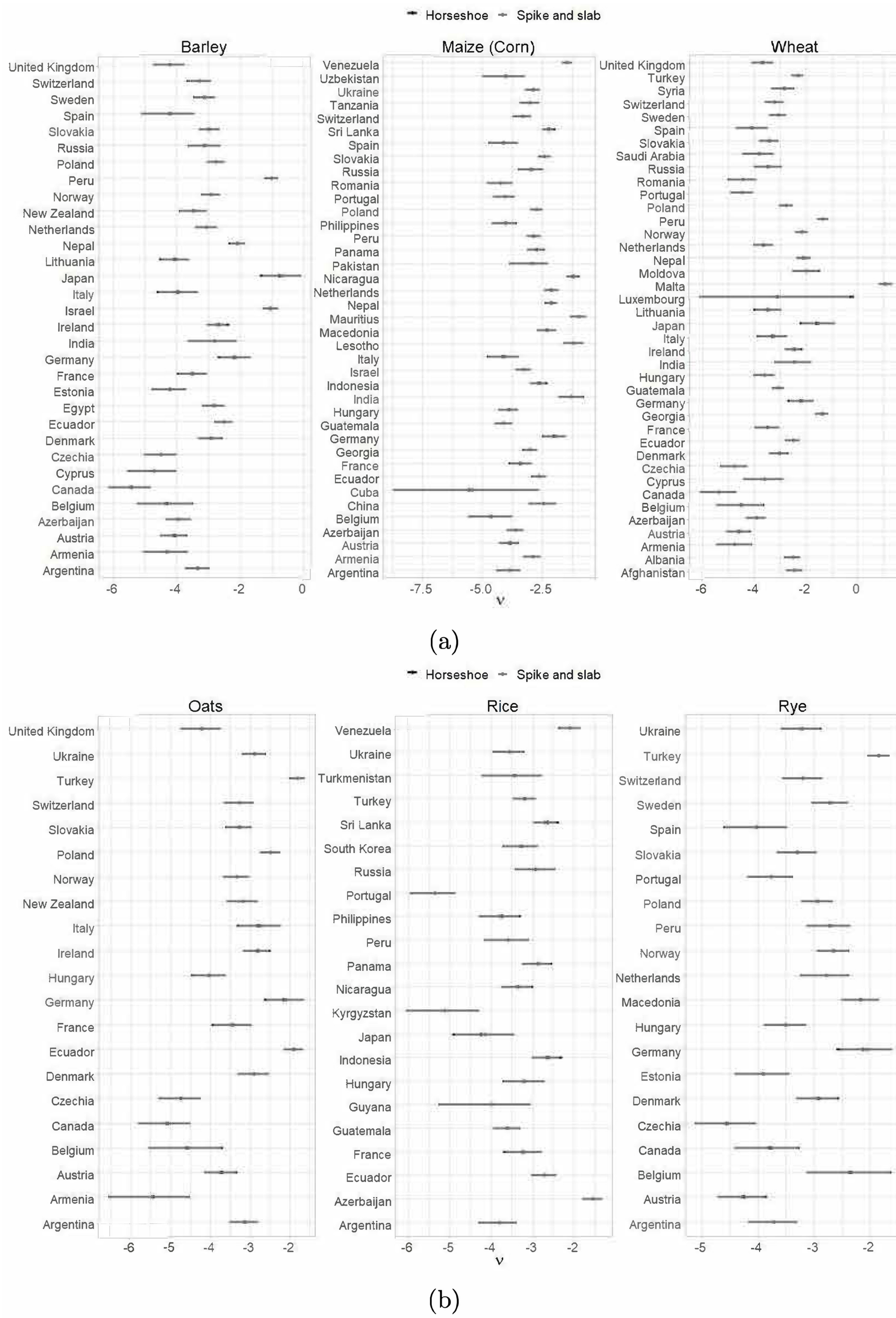


Fig. 15. Comparison between M1 and M2 for both point and interval estimates of country-crop random effects.

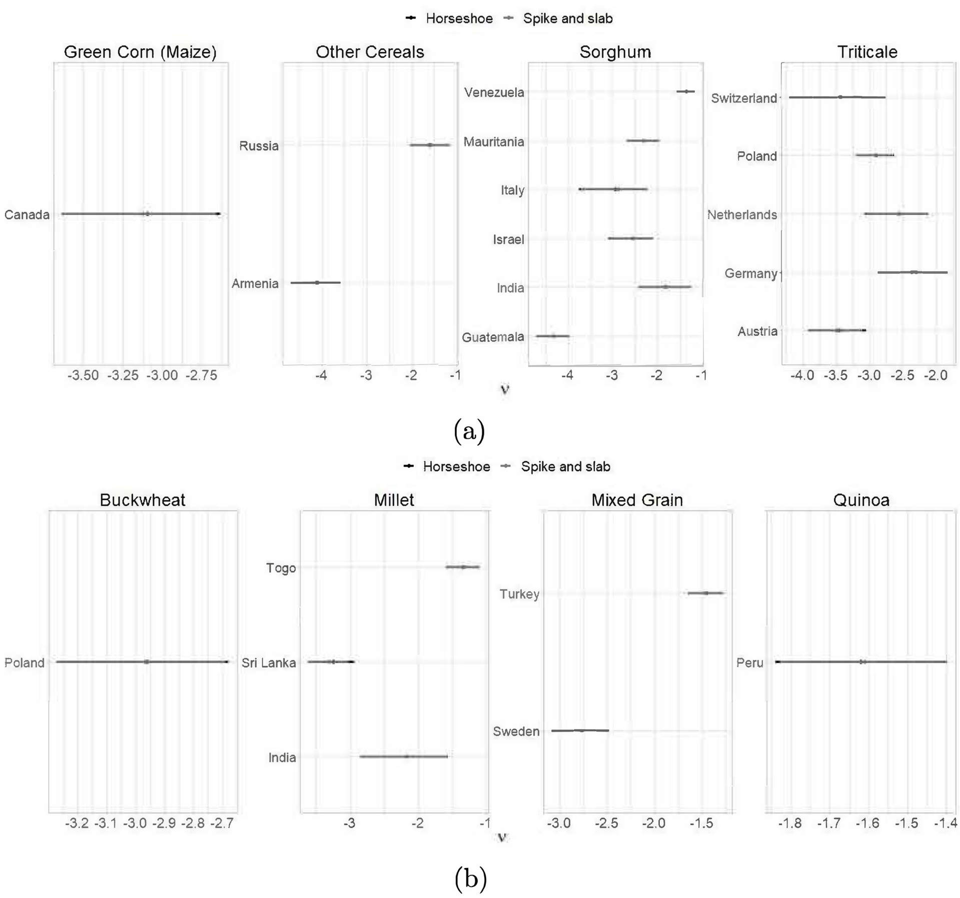


Fig. 16. M1 and M2 random effects point estimates and 95% credible intervals foreach cereal commodity (a) and (b).

7. References

Barbieri, M.M., and J.O. Berger. 2004. "Optimal predictive model selection." *The annals of statistics*, 32(3):870–897. DOI: https://doi.org/10.1214/009053604000000238.

Bhadra, A., J. Datta, N.G. Polson, and B. Willard. 2019a. "Lasso meets horseshoe: A survey." *Statistical Science*, 34(3):405–427. DOI: https://doi.org/10.1214/19-STS700.

Bhadra, A, J. Datta, N.G. Polson, and B.T. Willard. 2019b. "The horseshoe-like regularization for feature subset selection." *Sankhya B*. DOI: https://doi.org/10.1007/s13571-019-00217-7.

Boxall, R.A. 1986. "A critical review of the methodology for assessing farm-level grain losses after harvest," (working paper, Natural Resources Institute, University of Greenwich). Available at: http://gala.gre.ac.uk/10793/1/Doc-0235.pdf (accessed January 2020).

Breiman, L. 2001. "Random forests." *Machine learning*, 45(1):5–32. DOI: https://doi.org/10.1023%2fA:1010933404324.

Carvalho, C.M., N.G. Polson, and J.G. Scott. 2009. "Handling sparsity via the horseshoe" *Artificial Intelligence and Statistics*: 73–80. Available at: http://proceedings.mlr.press/v5/carvalho09a.html (accessed January 2020).

Carvalho, C.M., N.G. Polson, and J.G. Scott. 2010. "The horseshoe estimator for sparse signals." *Biometrika*, 97(2):465–480. Available at: http://www.jstor.org/stable/25734098 (accessed January 2020).

Celeux, G., F. Forbes, C.P. Robert, and D.M. Titterington. 2006. "Deviance information criteria for missing data models." *Bayesian analysis*, 1(4):651–673. DOI: https://doi.org/10.1214/06-BA122.

Cihan, P. 2018. "A comparison of five methods for missing value imputation in data sets." *International Scientific and Vocational Studies Journal*, 2(2):80–85. Available at: https://dergipark.org.tr/en/download/article-file/614943 (accessed January 2020).

Datta, J., and J.K. Ghosh. 2015. "In search of optimal objective priors for model selection and estimation." *Current Trends in Bayesian Methodology with Applications*, 225. DOI: https://doi.org/10.1201/b18502-12.

Fabi, C., A. English, M. Mingione, and G. Jona Lasinio. 2018. "Global foodloss index." FAO, Rome, Sdg 12.3. 1. Available at: http://www.fao.org/3/CA2640EN/ca2640en.pdf (accessed May 2020).

Ferrari, S., and F. Cribari-Neto. 2004. "Beta regression for modelling rates and proportions." *Journal of Applied Statistics*, 31(7):799–815. DOI: https://doi.org/10.1080/0266476042000214501.

FAO (Food and Agriculture Organization). 1980. *Assessment and collection of data on post-harvest food-grain losses*. Available at: http://www.fao.org/3/ca6157en/CA6157EN.pdf (Accessed January 2020).

FAO (Food and Agriculture Organization). 2014. SAVE FOOD: *Global Initiative on Food Loss and Waste Reduction. Definitional framework of food loss.* Available at: http://www.fao.org/3/a-at144e.pdf, (accessed January 2020).

FAO (Food and Agriculture Organization). 2019. *Moving forward on food losses and waste reduction.* Available at: http://www.fao.org/3/ca6030en/ca6030en.pdf (accessed January 2020).

FAOSTAT (Food and Agricultural Organization Statistics Division). 2016. *The United Nations Food and Agriculture Organization Database.* Database, 2016. Available at: http://faostat3.fao.org/home/E (accessed February 2021).

Franzin, A., F. Sambo, and B. Di Camillo. 2016. "bnstruct: an r package for bayesian network structure learning in the presence of missing data." *Bioinformatics*, 33(8):1250–1252. DOI: https://doi.org/10.1093/bioinformatics/btw807.

Gelman, A., J. Hwang, and A. Vehtari. 2014. "Understanding predictive information criteria for bayesian models." *Statistics and computing*, 24(6):997–1016. DOI: https://doi.org/10.1007/s11222-013-9416-2.

George, E.I., and R.E. McCulloch. 1993. "Variable selection via Gibbs sampling." *Journal of the American Statistical Association*, 88(423):881–889. DOI: https://doi.org/10.1080/01621459.1993.10476353.

Groll, A., J. Hambuckers, T. Kneib, and N. Umlauf. 2019. "Lasso-type penalization in the framework of generalized additive models for location, scale and shape."

Computational Statistics & Data Analysis, 140:59–73. DOI: https://doi.org/10.1016/j.csda.2019.06.005.

Grünberger, K. (2013). *Imputation of Loss Ratios for SUA/FBS Commodities*. A Technical Report produced as part of the Food and Agriculture Organization of the United Nations CapEx Project. Available at: http://www.fao.org/3/i4315e/I4315E.pdf (accessed February 2021).

Gustavsson, J., C. Cederberg, U. Sonesson, R. van Otterdijk, and A. Meybeck. 2011. *Global food losses and food waste.* Technical report, Food and Agricultural Organization (FAO). Available at: http://www.fao.org/3/a-i2697e/index.html (accessed July 2020).

Harris, K.L., and C.J. Lindblad. 1978. *Postharvest Grain Loss Assessment Methods*. Available at: http://pdf.usaid.gov/pdf_docs/PNAAG842.pdf (accessed July 2020).

International Energy Agency (IEA). 2019. *Statistics resources – balance definitions.* Available at: https://www.iea.org/statistics/resources/balancedefinitions/ (accessed November 2019).

Ijarchelo, S.M., K. Afereydoon, and L. Zamanzadeh. 2016. "Bayesian variable selection under collinearity of parameters." *Reserach journal of applied sciences*, 11:428–438. DOI: https://doi.org/10.3923/rjasci.2016.428.438.

Kuiper, M., and H.D. Cui. 2020. "Using food loss reduction to reach food security and environmental objectives – a search for promising leverage points." *Food Policy*: 101915. DOI: https://doi.org/10.1080/01621459.1988.10478694.

Kuo, L., and B. Mallick. 1998. "Variable selection for regression models." *Sankhyā: The Indian Journal of Statistics, Series B*, 65–81. Available at: https://www.jstor.org/stable/pdf/25053023.pdf?refreqid=excelsior%3A325e1cb3fa930037a8f90bdab30b9ac5 (accessed February 2021).

Mitchell, T.J., and J.J. Beauchamp. 1988. "Bayesian variable selection in linear regression." *Journal of the American Statistical Association*, 83(404):1023–1032. DOI: https://doi.org/10.1080/01621459.1988.10478694.

O'Hara, R.B., and M. J Sillanpää. 2009. "A review of Bayesian variable selection methods: what, how and which." *Bayesian analysis*, 4(1):85–117. DOI: https://doi.org/10.1214/09-BA403.

Plummer. M. 2003. "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling." In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March 20–22, 2003, Vienna, Austria. 2003. DOI: https://www.r-project.org/conferences/DSC-2003/Proceedings/.

Polson, N.G., and J.G. Scott. 2010. "Shrink lobally, act locally: Sparse bayesian regularization and prediction." *Bayesian statistics*, 9(501-538):105. DOI: https://doi.org/10.1093/acprof:oso/9780199694587.003.0017.

Polson, N.G., and J. G Scott. 2012. "Local shrinkage rules, lèvy processes and regularized regression." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2):287–311. DOI: https://doi.org/10.1111/j.1467-9868.2011.01015.x.

Sinharay, S., H.S. Stern, and D. Russell. 2001. "The use of multiple imputation for the analysis of missing data." *Psychological methods*, 6(4):317. DOI: https://doi.org/10.1037/1082-989X.6.4.317.

Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A. van der Linde. 2014. "The deviance information criterion: 12 years on." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493. DOI: https://doi.org/10.1111/rssb.12062.

Stekhoven, D.J., and P. Bühlmann. "Missforest–non-parametric missing value imputation for mixed-type data." *Bioinformatics*, 28(1):112–118. DOI: http://dx.doi.org/10.1093/bioinformatics/btr597.

Terenin, A.A., S. Dong, and D. Draper. 2019. "GPU-accelerated Gibbs sampling: a case study of the horseshoe probit model." *Statistics and Computing*, 29(2):301–310. DOI: https://doi.org/10.1007/s11222-018-9809-3.

Tibshirani, R. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288. Available at: www.jstor.org/stable/2346178 (accessed July 2020).

Transforming our world: the 2030 Agenda for Sustainable Development. 2015. Avaliable at: https://sdgs.un.org/2030agenda (accessed February 2021).

Vehtari, A., A. Gelman, and J. Gabry. 2017. "Practical Bayesian model evaluation using leaveone-out cross-validation and waic." *Statistics and computing*, 27(5):1413–1432. DOI: https://doi.org/10.1007/s11222-016-9696-4.

Waljee, A.K., A. Mukherjee, A.G. Singal, Y. Zhang, J. Warren, U. Balis, J. Marrero, J. Zhu, and P.DR. Higgins. 2013. "Comparison imputation methods for missing laboratory data in medicine." *BMJ open*, 3(8). DOI: https://doi.org/10.1136/bmjopen-2013-002847.

Watanabe, S. 2010. "Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory." *Journal of Machine Learning Research*, 11:3571–3594. Avaliable at: http://jmlr.org/papers/v11/watanabe10a.html (accessed July 2020).

White, I.R., P. Royston, and A.M. Wood. 2011. "Multiple imputation using chained equations: issues and guidance for practice." *Statistics in medicine*, 30(4):377–399. DOI: https://doi.org/10.1002/sim.4067.

Wu, L. 2009. Mixed Effects Models for Complex Data. CRC Press.

Received February 2020
Revised July 2020
Accepted November 2020