Environmental Development 54 (2025) 101141

Contents lists available at ScienceDirect

Environmental Development

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

Herding resilience: Surveys and Bayesian spatial models for Africa's livestock

Tamás Krisztin [a,*,1], Michiel van Dijk [b,a], Philipp Piribauer [c]

[a] *International Institute for Applied Systems Analysis, Laxenburg, Austria*
[b] *Wageningen Economic Research, The Hague, Netherlands*
[c] *Austrian Institute for Economic Research, Vienna, Austria*

ARTICLE INFO

JEL classification:
Q12
Q18
Q54
C11
O13

Keywords:
Livestock distribution
INLA models
Climate resilience
Household surveys
African agriculture

ABSTRACT

This paper proposes a novel method for mapping livestock distribution in Africa using the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA). Using a Bayesian spatial statistical model, we produce maps of livestock distribution at a resolution of 1 km^2. Our case study in Malawi, covering 2010 and 2019, demonstrates the effectiveness of the method in mapping five livestock species. We compare our results with the Gridded Livestock of the World (GLW) database and use the maps to assess livestock vulnerability to climate-related flood risks under different climate scenarios. This approach provides a rapid, data-rich tool for policy makers to assess climate risks to livestock, which is critical for sustainable agricultural development and environmental management in data-poor regions.

1. Introduction

Livestock plays a central role in African agriculture, significantly shaping land use, greenhouse gas emissions and dietary patterns. Livestock products (meat, milk and eggs) contribute to 15% and 31% of the global per capita calorie and protein supply, respectively (Godde et al., 2021). This global perspective underscores the importance of livestock in dietary patterns, which is particularly relevant for Africa, where livestock is critical to nutritional outcomes and food security. Approximately 30% of global ruminant meat and 6% of milk production comes from grazing systems on land that is often unsuitable for crop production, a situation common in several parts of Africa where pastoralism and extensive grazing systems are prevalent. Recent studies indicate that climate change poses a serious threat to the supply of livestock products. It has been estimated that 4%–19% of livestock production areas are facing an increased risk to heat stress by 2100 (Rahimi et al., 2021).

Understanding the distribution and impact of livestock is critical to addressing food security, environmental sustainability, climate and health challenges. Livestock's contribution to agricultural land use, its role in global greenhouse gas emissions and its role in human diets underline its importance. The transmission of diseases from livestock to humans, as demonstrated by the avian influenza outbreak in 2007, further underlines the need for comprehensive livestock management strategies (Uwizeye et al., 2020; Chiriacò and Valentini, 2020; Eisen and Brown, 2021; Santeramo et al., 2020).

Accurate and detailed mapping of livestock distribution is essential for a range of scientific and practical applications, including climate change research, epidemiological research and environmental management. The effects of climate change are not evenly

* Corresponding author.
E-mail addresses: krisztin@iiasa.ac.at (T. Krisztin), michiel.vandijk@wur.nl (M. van Dijk), philipp.piribauer@wifo.ac.at (P. Piribauer).

[1] The research was supported by funds of the Austrian National Bank (OeNB) (18307).

https://doi.org/10.1016/j.envdev.2025.101141
Received 29 March 2024; Received in revised form 18 December 2024; Accepted 19 January 2025
Available online 27 January 2025

distributed and will have different impacts in terms of heat stress (Godde et al., 2021; North et al., 2023) or flooding (Palmer et al., 2023) across the African subcontinent. Having accurate, high-resolution maps of the current location of livestock would be crucial for both international and national policy makers to effectively target high-risk areas.

Unfortunately, the global availability of spatially explicit information on livestock distribution is somewhat limited. The granularity and accuracy of livestock distribution data varies considerably between regions, with well-documented and high-resolution datasets available mainly in developed countries. In contrast, regions with limited resources, such as the African subcontinent, often suffer from a lack of spatially explicit data. This discrepancy poses significant challenges for conducting comprehensive global and national assessments, and formulating effective policies for disease control, environmental protection and sustainable agricultural development.

The challenges of producing accurate, high-resolution maps of livestock distribution are manifold. In areas where data is scarce or non-existent, traditional methods of data collection are often impractical or cost prohibitive. In addition, the dynamic nature of livestock populations, influenced by seasonal migration, trade and changes in agricultural practices, further complicates the task of maintaining up-to-date and accurate distribution maps. This lack of accurate data hampers our ability to assess the environmental impact of livestock, understand the transmission dynamics of zoonotic diseases, and target interventions to regions of greatest need.

A key resource in this area is the Gridded Livestock of the World (GLW) database, which provides a global perspective on the distribution of four major livestock species. This resource has been developed through the efforts documented in Wint and Robinson (2007) and Robinson et al. (2014). While the GLW provides a basic global overview, the need for more detailed regional data has driven additional research. Efforts to produce more specific maps for particular regions and animal species have been reported in studies such as Neumann et al. (2009), Van Boeckel et al. (2011), Prosser et al. (2011) and Van Boeckel et al. (2012). In addition, the mapping of livestock production systems, which is essential for a comprehensive understanding of agricultural practices, has been explored in works such as Robinson et al. (2011) and Gilbert et al. (2015). Collectively, these studies improve our understanding of livestock distribution and are crucial for advancing related scientific and practical applications.

A recurring problem in livestock mapping is the limited availability of data and the need to extrapolate from patchy data that is available only for a selected number of sub-national regions within a country. Most work has used data from provincial or district level census along with Ordinary Least Squares to extrapolate information. Recent work has used advanced machine learning techniques to create a global livestock map (Nicolas et al., 2016; Gilbert et al., 2018) or spatial statistical methods for modeling urban expansion (Krisztin et al., 2022) or disease mapping in Africa (Moraga et al., 2021).

To the best of our knowledge, there are no studies using nationally representative household surveys to produce livestock maps. A growing number of detailed agricultural household surveys are available that provide detailed information on farmers' livestock holdings. For Africa, the most relevant is the World Bank-supported Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA). Such surveys provide rich data sources with geocoded information for countries in sub-Saharan Africa. This has been used in several recent studies on agricultural and food security issues (see e.g. Julien et al. (2019), Van Dijk et al. (2020)). The LSMS-ISA surveys provide a valuable and readily available source of data for mapping livestock distribution. Unlike global livestock distribution products, which take considerable time to compile, these surveys are regularly updated and are publicly available, and therefore can be directly used as input for livestock mapping. Consequently, when used effectively, they provide an important resource to inform livestock policies in data-poor regions of Africa.

The main contribution of this paper is to propose a method for reconstructing livestock distributions from LSMS-ISA surveys. We use a Bayesian spatial statistical model where spatial effects are modeled using stochastic partial differential equations. The main advantage of our approach are in the high-resolution maps (1 km^2) that the empirical method can readily produce from geolocated household surveys. Another key feature is that due to the Bayesian estimation the model provides clear uncertainty estimates. This is a useful tool for policy makers to assess climate risks to livestock relatively quickly.

Our approach is demonstrated using a case study in Malawi to map the location of five types of livestock in 2010 and 2019. Malawi is a useful case study because the livestock sector is an important contributor to economic growth and household income in Malawi. In 2019, it will contribute about 7% to national GDP and about 29.5% to agricultural GDP (Food and Agriculture Organization of the United Nations, 2022). Livestock was also estimated to contribute between 19 and 54% to total household income (Kaumbata et al., 2020). In Malawi, we compare our results with the 2010 GLW database. As flooding is considered a major climate change challenge in Malawi, we also illustrate our approach to identifying the type and number of livestock most vulnerable to flood-related climate extremes by overlaying the livestock distribution maps with maps that show areas at high risk of being flood under a high-emissions scenario.

The remainder of the paper is structured as follows. Section 2 presents the empirical specification. Section 3 discusses both the household survey data and the spatial factors considered. Section 4 presents the predicted livestock distribution in 2010 and 2019, the comparison of the predicted with a global livestock dataset, and an assessment of livestock flood risk in Malawi. Section 5 discusses our findings, while the last section concludes.

2. Empirical specification

Our modeling framework is designed to predict livestock counts across a spatial domain where direct survey information is unavailable. We therefore employ a hierarchical geostatistical negative binomial model to account for the count nature of the data and the presence of overdispersion, which is typical in livestock distributions.[2] Let y_i represent the number of livestock in location

[2] As a robustness check we have also considered the Poisson, but the negative binomial can handle additional variance, which resulted in a clear under-performance relative to the negative binomial specification.

s_i for $i = 1, \ldots, n$. The negative binomial distribution appears particularly appealing since it generalizes the Poisson distribution by accounting for overdispersion, which allows the variance to exceed the mean.[3]

We therefore consider a latent Gaussian model where the conditional expectation of y_i, given a set of explanatory variables $\boldsymbol{x}_i$, is linked to a structured additive linear predictor $\boldsymbol{x}_i\boldsymbol{\beta}$ through a log-link function. The first two moments can be written as:

$$E(y_i|\boldsymbol{x}_i) = \log(\lambda_i) = \boldsymbol{x}_i\boldsymbol{\beta} + \xi_i \tag{1}$$

$$Var(y_i|\boldsymbol{x}_i) = \lambda_i\left(1 + \frac{\lambda_i}{\alpha}\right), \tag{2}$$

where $\boldsymbol{x}_i$ denotes an $1 \times q$ vector of explanatory variables with corresponding $q \times 1$ vector of slope coefficients $\boldsymbol{\beta}$. The explanatory variables contain spatial predictors of livestock from the categories (motivated by Gilbert et al. (2018)) land use, anthropogenic causes, vegetation and satellite data, climatic condition, environmental conditions and topographic determinants. The parameter α is an overdispersion parameter, which allows to accommodate extra variability compared to a Poisson specification. When α becomes large, the negative binomial distribution approaches the Poisson.

Our modeling framework also allows to explicitly control for spatial random effects. We therefore relate the point-based farm-level survey data to a continuous representation of spatial effects. The spatial structure in our model is represented by ξ_i, which is modeled as a Gaussian process. This latent process introduces correlation between locations based on their spatial proximity. The $n \times 1$ vector $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_n]'$ follows a zero-mean Gaussian process with a Matérn covariance structure $\mathcal{M}(d)$,

$$\text{Cov}(\boldsymbol{\xi}) = \sigma_\xi^2 \mathcal{M}(d) \tag{3}$$

$$\mathcal{M}(d) = \frac{1}{\Gamma(\upsilon)2^{\upsilon-1}}(kd)^\upsilon K_\upsilon(kd), \tag{4}$$

where σ_ξ^2 is the marginal variance. k and υ are non-negative parameters of the covariance function and K_υ denotes the modified Bessel function of second kind and order υ. The spatial dependence is determined by the Euclidean distance between two locations, denoted by $d = \|s_i - s_j\|$. Since the covariance between two locations depends only on the Euclidean distance (and not on the direction), the process is said to be isotropic (see Cressie and Wikle (2015)).

Such an approach models livestock as a spatial smooth process. Livestock distributions are most likely not smooth in space as the location of livestock is strongly determined by land cover (e.g. grassland), population density and other spatial drivers. However, if one accounts for these drivers, which is done in the INLA framework through the inclusion of a high number of fixed effects, which capture the spatial discontinuities, the assumption of a smooth spatial process to model the probability of livestock location is realistic.

The hierarchical model specification features a decomposition of the error term into a spatially unstructured term and a spatial random effect. The spatial random effect is defined continuously over space and provides information on spatial discontinuities. A particular advantage of the continuous spatial representation over spatial econometric models for areal data (see LeSage and Pace (2009)) is that it avoids problems of data aggregation into discrete areal representations and pre-defining spatial neighborhood structures. Although recent advances in spatial econometric methods increasingly use estimated spatial neighborhood schemes (see, for example, Lam and Souza (2020), Krisztin and Piribauer (2023), or Piribauer et al. (2023)), a (continuous) spatial model specification sketched above avoids problems of spatial discontinuities much better than model specifications for gridded data, and therefore appears to be superior in terms of revealing an underlying data generating process that is continuous in space (see, for example, Simpson et al. (2012), or Laurini (2017)).

To reduce the resulting computational burden, the influential work of Lindgren et al. (2011) advocates using a discrete Gaussian Markov random field (GMRF) representation of the (continuous) Gaussian field. The GMRF representation allows the use of computationally efficient (matrix) operations, since GMRF representations are characterized by sparse precision matrices (for a detailed discussion see Rue and Held (2005) and Rue et al. (2009a)). The discrete representation as a GMRF also allows the use of the Bayesian Integrated Nested Laplace Approximation (INLA) approach (Rue et al., 2009b) as a computationally efficient alternative to Markov-chain- Monte Carlo (MCMC) sampling. INLA bypasses time-consuming MCMC algorithms by performing direct numerical approximations to the posterior distributions of interest. Several studies have also shown that the INLA approach to estimation and inference leads to particularly accurate posterior parameter distributions and avoids potential mixing problems of MCMC sampling (Blangiardo and Cameletti 2015, Lindgren and Rue 2015).

3. Data and projections

3.1. Livestock data

The World Bank, in collaboration with national statistical offices, has supported the implementation of nationally representative household surveys with a strong focus on agriculture through the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) project. Survey results are available for eight sub-Saharan countries, covering multiple years. The LSMS-ISA collects information on a wide range of agricultural variables, including information on the ownership of the main types of livestock. It also includes the geo-coordinates of all enumeration areas, allowing the information to be mapped and combined with other spatially explicit data. The geo-coordinates are presented with a random offset of 0–10 km to preserve the confidentiality of sample household. The offset points are constrained at the district level, so that they still fall within the correct administrative unit and therefore do not affect district-level figures for livestock totals.

[3] Note that the Poisson specification assumes that its mean equals the variance.

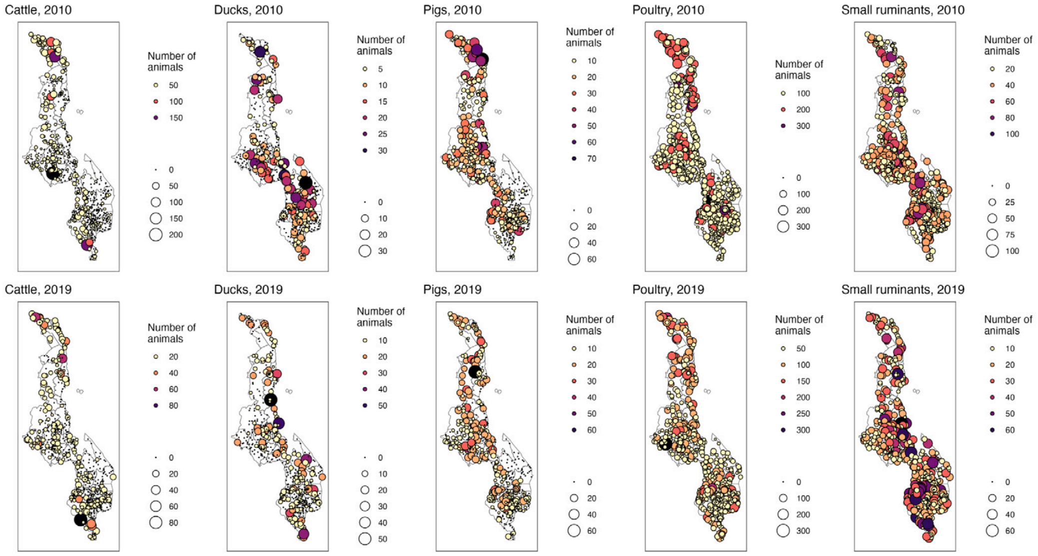

Fig. 1. Distribution of animals in IHS3 and IHS5 surveys. The colors and sizes of the circles both represent the number of animals reported. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We use Malawi as a case study. The main data sources are the Third Integrated Household Survey (IHS3) and the Fifth Integrated Household Survey (IHS5), which are representative at the district, regional, urban-rural and national levels, covering 12,271 and 12,288 households respectively in 768 enumeration areas. Farmers were asked to report whether they owned different types of livestock (e.g. cows, goats, ducks and poultry) and how many they owned in 2010–2011 and 2019–2020. We reclassified the different types of livestock into five categories that are also used by Robinson et al. (2014) for global livestock mapping: cattle, dugs, pigs, poultry, and small ruminants (goats and sheep). The distribution of livestock within Malawi and its administrative regions is shown in Fig. 1.

3.2. Determinants of location

Wint and Robinson (2007) and Robinson et al. (2014) used a comprehensive suite of gridded datasets as covariate variables to predict the distribution and abundance of global livestock populations. Their analyses predominantly included variables within five categories: land use, anthropogenic determinants, vegetation, climate, environmental conditions and topography. In our study, we take a similar approach by integrating spatial predictors from these established categories into our model. However, we place greater emphasis on high spatial resolution and focus our efforts on sub-Saharan Africa. To improve the predictive power of our model in this specific context, we have introduced two additional variables: tsetse fly species diversity, which is a critical factor influencing livestock dynamics in Africa (Ekwem et al., 2021), and a road buffer to represent accessibility. Table 1 summarizes the gridded layers used in our analysis, including these new additions.

3.3. Projections

To produce projections of livestock numbers across Malawi, an ex-post allocation procedure was implemented for predictions derived from the Integrated Nested Laplace Approximation (INLA) model. This post-processing step harmonizes model projections with national livestock totals reported by the Food and Agriculture Organization (FAO) for 2010 and 2019. Livestock densities within each of Malawi's administrative districts were calculated by allocating predictions from the Integrated Nested Laplace Approximation (INLA) model. This allocation was done to match the national total livestock numbers from FAO animal stock totals for 2010 and 2019, respectively.[4]

The LSMS-ISA household survey is representative at the administrative regional level. For this reason, the national total livestock counts were in a first step disaggregated to the subnational regions to provide the highest resolution of accuracy from the LSMS-ISA survey. This was achieved by aggregating the relative proportions of subnational shares by administrative region derived from the LSMS-ISA household survey.

[4] For ducks, FAO does not report official statistics for Malawi. Here we used the 2010 GLW and scaled it up by the growth rate of duck stocks in the period 2010 to 2019, as reported by FAO for the Southern Africa supranational region.

Table 1
Table of spatial determinants.

Type	Variables	Use	Source
Land use	IUCN world database of protected areas, available in 2010 and 2019	Mask	IUCN et al. (2010)
Anthropogenic	Human population density in people per km^2, available in 2010 and 2019 (Worldpop)	Spatial predictor and mask	Tatem (2017)
	Travel time in minutes to cities of 50,000 people	Spatial predictor	Weiss et al. (2018)
	Road buffer (25 km), based on the Global Roads Inventory Project (GRIP)	Spatial predictor	Meijer et al. (2018)
Vegetation	Cropland cover, classes 10, 11, 12, 20, 30, 40; available in 2010 and 2019 (ESA-CCI)	Spatial predictor	Defourny et al. (2012)
	Grassland cover, classes 110, 130; available in 2010 and 2019 (ESA-CCI)	Spatial predictor	
	Settlement cover, class 190; available in 2010 and 2019 (ESA-CCI)	Spatial predictor	
	Forest cover, classes 50, 60, 61, 62, 70, 71, 72, 80, 81, 82, 90, 100; available in 2010 and 2019 (ESA-CCI)	Spatial predictor	
Climatic	Global aridity index and potential evapotranspiration	Spatial predictor	Zomer et al. (2022)
Environmental	Number of Tsetse fly species, (FAO)	Spatial predictor	Cecchi et al. (2015)
Topography	Average elevation in meters	Spatial predictor	NASA JPL (2013)

The INLA projections for each administrative region are subsequently distributed based on the INLA model projections, taking into account three masking data sets. First, adjustments to the allocated livestock counts were made using a suitability mask (following Gilbert et al. (2018)), which strictly excludes areas deemed unsuitable for livestock. Specifically, this mask excludes areas predominantly covered by water (more than 50% of an area, as detailed in Table 1).[5] Second, the masking excludes areas with human population densities exceeding 10,000 individuals per km^2 in Asia and Africa, following the criteria set by the human population data layer and as discussed in Gilbert et al. (2018). Third, a global exclusivity mask was created based on the 2010 iteration of the World Database of Protected Areas, with IUCN categories Ia, Ib, II and III designated as unsuitable. These categories are characterized by strong conservation efforts and significant restrictions on human activities, making them less susceptible to livestock encroachment and other grazing practices.

Given these masking areas, projections from the INLA model are scaled to integrate over all pixels within an administrative region, ensuring consistency with total livestock counts as derived from FAO and the LSMS-ISA survey. To mitigate overestimation, a predetermined maximum livestock number[6] from the survey is applied as a cut-off threshold. Although this approach improves aggregate accuracy, it also introduces potential limitations. Specifically, the fixed cut-off, although effective in excluding extreme outliers, may obscure genuine spatial heterogeneity in regions where higher livestock densities are plausible but unobserved. Notably, this threshold was exceeded only in projections for ducks (for < 2% of pixels); all other livestock categories showed no instances of exceeding the cut-off across the spatial domain. The observed discrepancies in duck projections may be attributed to the relatively poor model fit for this category, likely driven by the limited proportion of non-zero observations (27%).

The INLA estimation method employed in this analysis provides a Gaussian approximation to the marginal posterior distributions, along with default approximations for the 2.5th and 97.5th quantiles (additional quantiles can be obtained as needed).[7] This approach facilitates straightforward uncertainty quantification for the estimates. To compute uncertainty bounds, we apply the same masking and projection method described above to the 2.5th and 97.5th quantiles.

4. Results

4.1. Spatial model

We ran a spatial simulation using the INLA model package in R,[8] using default priors for the linear parameters. For the spatial model, we constructed a mesh with a maximum triangle edge length for the inner and outer boundaries of 8 km and 60 km respectively. Estimation is based on INLA using the approach proposed in Lindgren et al. (2011), where the default zero mean and 0.1 precision prior was used. Separate models were estimated for each of the five livestock types and for each survey periods of LSMS-ISA (2010 and 2019).[9] The linear coefficient estimates reveal several patterns across livestock types and years. Population density consistently shows a significant negative relationship at the 95% confidence interval in multiple models. Notably, the number of tsetse species is negatively correlated with livestock but positively correlated with pigs. Additionally, pigs are more likely to be

[5] Note, that for the study this only affect pixels on the coast of lake Malawi.

[6] Three times the maximum number of livestock reported in the survey.

[7] Note that applying the masking and projection method directly to INLA-based posterior quantiles does not yield the quantiles of the transformed distribution, which may introduce uncertainty quantification inaccuracies.

[8] https://www.r-inla.org

[9] We have decided to estimate the two time points separately to allow for greater flexibility in capturing potential heterogeneity in the parameters across the two periods. In a pooled model, fixed effects can account for time-invariant characteristics, but they may not fully capture structural differences in livestock distribution processes over time.

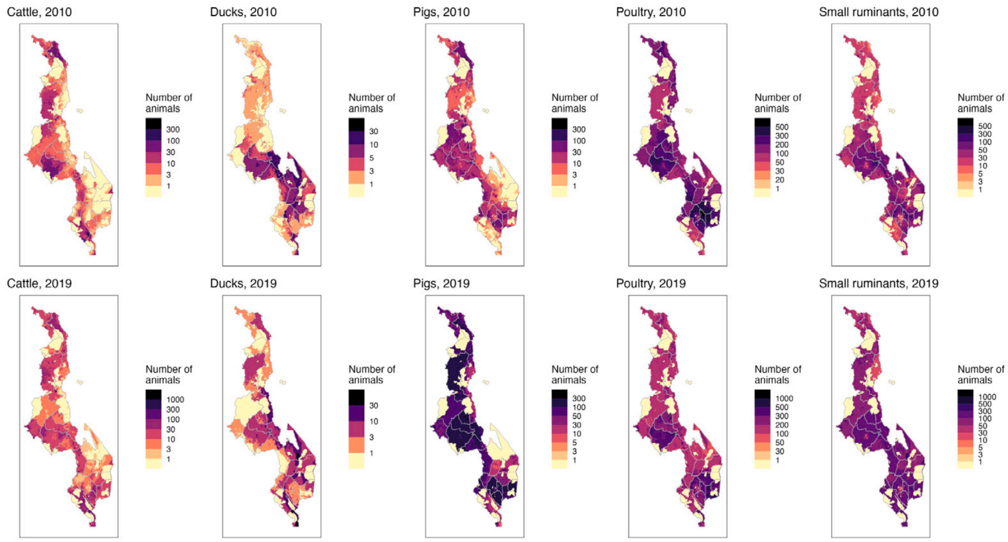

Fig. 2. Predicted livestock distribution in 2010 and 2019.

located closer to roads. For further details on the linear coefficient estimates and the prior specification, refer to the Supplementary Information (SI).

The posterior means for all livestock types and years are shown in Fig. 2; lower 2.5th and upper 97.5th quantile livestock distributions (which provide an assessment of prediction uncertainty) are presented in the SI. The plots illustrate the distribution of livestock in Malawi in 2010 and 2019, focusing on five livestock types: cattle, ducks, pigs, poultry and small ruminants. In 2010, the distribution patterns show varying densities of livestock across the categories, with cattle and small ruminants showing the highest numbers, reaching up to 300 individuals in some locations. Pigs and poultry also have a significant presence, but with lower peak numbers, suggesting a diversified but unevenly distributed livestock sector across the country. Note the strong concentration of poultry and small ruminants around Lilongwe and Blantyre, while cattle are more dispersed.

By 2019 there are notable changes in the distribution patterns of livestock types. There is an increase in the number of poultry, with numbers peaking at 500 individuals in some areas, indicating a significant shift. Conversely, the distribution of pigs and small ruminants shows a decrease in the highest numbers observed. The cattle and duck categories remain relatively stable, with slight variations in their distributions.

These changes suggest a change in farming practices and management, possibly caused by agricultural policy interventions over the period 2010–2019. Over this period, the government in Malawi introduced several plans and strategies to support the development and diversification of the agricultural sector, including the Farm Income Diversification Program (2010–2014), the National Agricultural Plan (2016–2021) and the National Agricultural Investment Plan (2017–2022) (Food and Agriculture Organization of the United Nations, 2022). Although, studies on the impact of these policies on livestock sector development are scant, there is anecdotal evidence that they might have resulted in diversification towards livestock production among beneficiary farmers, which potentially explains the observed change in livestock distribution patterns (Ng'ong'ola et al., 2017).

4.2. Comparison to GLW and validation

We compare our 2010 results with the GLW 2010 data (Robinson et al., 2014), which contains global livestock maps for cattle, ducks, pigs, poultry and small ruminants.[10] The spatial resolution of the GLW is coarser at 0.083333 decimal degrees (about 10 km at the equator). Note that the administrative level and national totals of the 2010 GLW product differ from the FAO totals, to which our predictions were re-scaled to. To facilitate a fair comparison, we re-scale the 2010 GLW data to match aggregate FAO statistics.

Fig. 3 shows a comparison of livestock distribution in Malawi for 2010 according to household survey results and GLW data. The circles on the maps indicate the locations of the LSMS household surveys, with different circle sizes representing the number of livestock owners reported. Across the five livestock types – cattle, ducks, pigs, poultry and small ruminants – there is a discernible pattern of discrepancies between the survey data and the GLW estimates.

For cattle, the survey points show a consistent under-representation in the GLW data, with differences of up to 400 in some areas. Similarly, for small ruminants, there is a significant underestimation by the GLW, particularly in regions where household

[10] The GLW database contains two versions: the areal-weighted livestock distribution, which distributes animal count data evenly over a census polygon, and the dasymetric distribution, which assigns animal densities to different pixels within a given census polygon according to a random forest model. We use the dasymetric method as a benchmark.

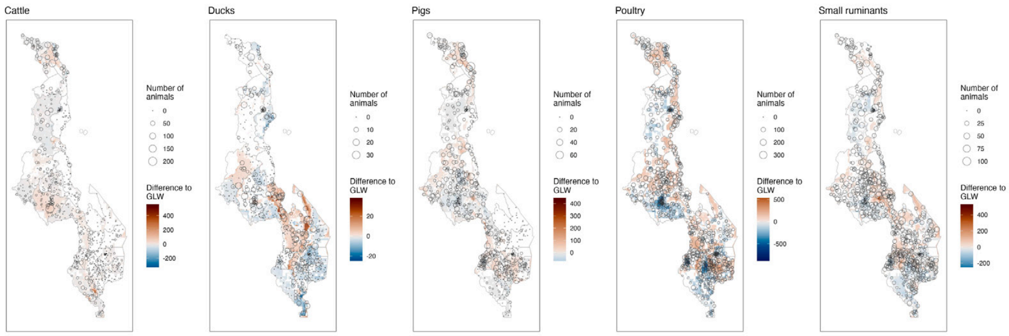

Fig. 3. Comparison of livestock distribution in 2010 with Gridded Livestock of the World (GLW). Positive numbers indicate under-, while negative numbers indicate over-representation of livestock in the GLW data, as compared to our predictions. Circles correspond to the location of LSMS households, with the size of the circle indicating the number of reported livestock heads. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

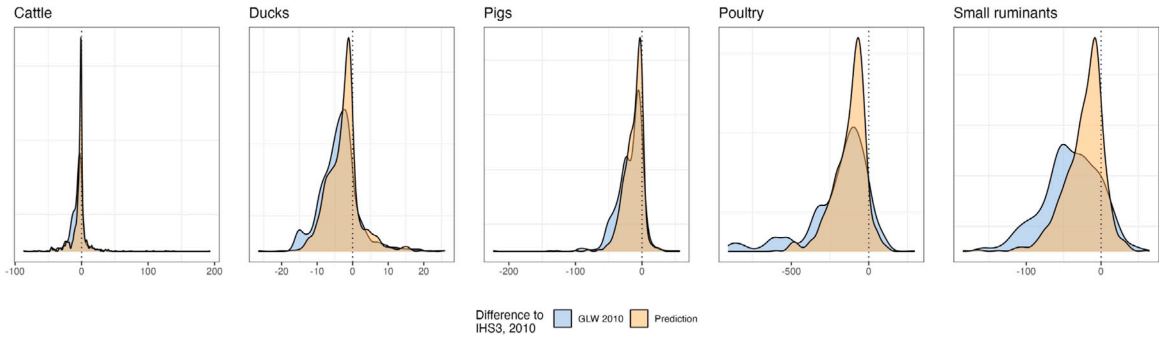

Fig. 4. Difference between the IHS3 data at the enumeration points (768 points) with either Gridded Livestock of the World (GLW) and the INLA predictions. The distributions of deviations are depicted, where positive values signify underestimation and negative values overestimation relative to the survey counts. The dotted vertical line represents the line of unity, indicating an exact match between survey data and predictions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

surveys report numbers of more than 100 animals. In contrast, the poultry data suggest an overestimation of the GLW data, with differences of up to −600 head in some regions. The distributions for ducks shows more modest variations, with discrepancies in the range of −20 to +20 head. In the case of pigs, GLW indicates a spatially rather more concentrated pig distribution, while our predictions are more wide spread spatially, which is generally supported by the LSMS survey points. These patterns suggest that the GLW data may not fully capture the reality on the ground as reported by the LSMS household surveys, particularly for cattle and small ruminants, and to a lesser extent for poultry, where the GLW appears to overestimate numbers.

Fig. 4 presents a comparison of livestock counts from enumeration points (768 points) of the IHS3 survey with either those reported in the 2010 GLW dataset and the outputs from our Bayesian predictions. Notably, GLW estimates consistently exceed the IHS3 observations across all categories of livestock. This could indicate a general trend of overestimation by GLW. This discrepancy is marginally observable in the cattle category but becomes significantly pronounced for poultry. In contrast, the INLA model deviations closely align with zero, suggesting negligible systematic bias and a robust congruence with the IHS3 survey data. The density of these deviations closer to the zero line underscores the INLA model's accuracy in representing actual livestock counts as per the survey.

The observed discrepancies highlight the need for more accurate data collection methods, such as household surveys, to improve the accuracy of livestock distribution models. The significant under-representation of cattle and small ruminants in the GLW data could have important implications for agricultural policy and climate resilience planning. The over-representation of poultry numbers in GLW is likely due to the large discrepancies in total numbers — GLW suggests around 15 million animals, while IHS3 suggests around 8 million. These findings highlight the importance of integrating localized survey data into global livestock distribution models to ensure more reliable data for policy makers and stakeholders working towards sustainable agricultural development and adaptation to climate change in Malawi.

In addition we conducted a cross-validation exercise to robustly assess the model's performance. Specifically, we estimated the model using 70% of the data points, reserving the remaining 30% for out-of-sample prediction. To ensure that the distribution of zeros was preserved, we applied a stratified random sampling approach. This prevented the validation dataset from being dominated by zero observations, which could bias the evaluation. The procedure was repeated over 200 iterations to account for variability

Table 2
Out-of-sample RMSE values for livestock categories with mean and standard deviation, as well as comparison to an intercept only benchmark model.

Livestock	Mean RMSE	SD RMSE	RMSE relative to benchmark
Cattle	6.12	1.35	0.9962
Ducks	6.15	1.32	0.9954
Pigs	5.64	0.55	0.9914
Poultry	27.29	2.56	0.9705
Small ruminants	10.32	0.44	0.9921

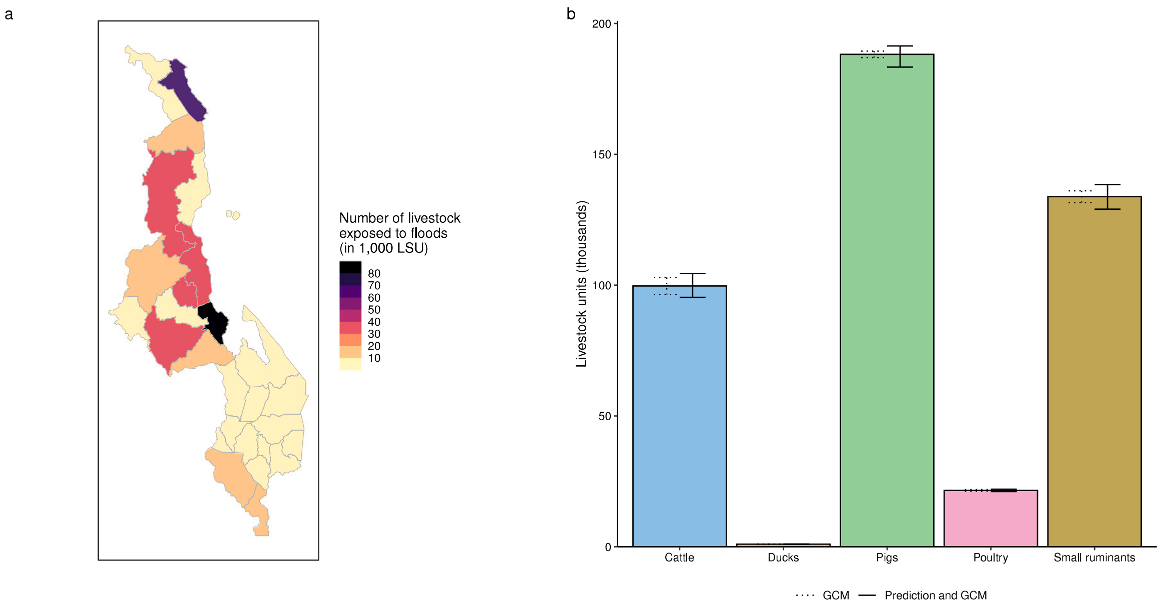

Fig. 5. Spatial distribution of livestock in 2019 under SSP2, RCP8.5 scenario, illustrating the long-term risk of 100-year riverine floods. Livestock units are aggregated based on the FAO standard, where cattle are considered as 0.5 LSU, ducks 0.01 LSU, pigs 0.2 LSU, poultry 0.01 LSU, and small ruminants 0.1 LSU. *Note: The dotted error bars represent uncertainty from the use of five global circulation models (NorESM1-M, GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM), while the solid error bars additionally include the uncertainties of the Bayesian estimation (95% confidence intervals)*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

introduced by the random selection of training and validation sets. The results of this exercise, summarized in Table 2, provide a detailed comparison of the models' out-of-sample predictions using root mean squared error (RMSE). The second column shows the mean RMSE over 200 iterations, while the third column contains the respective standard deviation of RMSE scores. This approach offers a robust evaluation of the model's ability to capture the data-generating process and provides insights into its accuracy and reliability. The final column presents the ratio of the mean RMSEs to an intercept only model (values below one indicate an improvement over the benchmark).

4.3. Flood risk

Malawi's agricultural sector, particularly smallholder farmers, is highly vulnerable to climate extremes. Of these, flooding is considered to be the most frequent and damaging climate event (World Bank Group, 2022). Since 2010, Malawi has experienced at least 16 major flood events, the most recent of which was in 2019 when Cyclone Idai hit Malawi. In addition to the devastating impact on infrastructure, housing and crops, livestock losses were estimated at around 8.2 million, or 3.7% of the total damage, and more than 47,000 animals were killed.[11]

This section illustrates how gridded, high-resolution livestock maps based on household surveys can be used to identify the type and number of livestock most vulnerable to climate extremes, particularly flooding.

To measure the risk of exposure to river flooding, we used data from the World Resources Institute's Aqueduct flood risk analyser (Ward et al., 2020), which presents flood risk maps for the high-emissions (referred to as RCP8.5; see Riahi et al. (2011)) global

[11] Source: https://reliefweb.int/.

warming scenario using six different global circulation models. For illustrative purposes, we selected the maps representing 100-year flood maps (i.e. floods occurring once in 100 years), which are also used as the reference for global flood risk studies (Tellman et al., 2021). We combined the flood maps with the different livestock maps to identify the number and location of livestock units at risk of flooding.

Fig. 5 shows the number of livestock exposed to flood risks in Malawi, measured in African Livestock Units (LSU). The figures provide a visualization of the differences in the exposure of different types of livestock to flood risks. For ease of comparison, all livestock have been converted to LSU. Panel a shows an aggregate map by district of the total LSU per district. The spatial distribution shows that the main exposure of livestock is in northern and central Malawi, close to the estuaries of Lake Malawi. Panel b compares the types of livestock (in LSU) exposed to 100-year flood events. Pigs have by far the highest exposure, with the number of livestock units at risk close to 200k LSU. This is a significant proportion of the livestock population, suggesting that pigs are the most vulnerable to flood-related losses within the agricultural sector. Small ruminants show a similar risk pattern due to their larger numbers in agricultural practice. Conversely, poultry, ducks and cattle have significantly lower exposure levels, ranging from 5,000 to just under 20,000 LSU. Note, that we consider two types of uncertainties with our flood risk estimates. First, the uncertainty of climate projections: we use five flood risk maps from different global circulation models (see the SI for the individual flood risk maps). Second, our proposed INLA estimation framework provides uncertainty about the livestock distribution, which can be easily factored in for more informed policy recommendations. The INLA estimation method provides 2.5th and 97.5th quantiles of the (approximate) posterior distribution. These are used in combination with the flood risk maps to calculate the upper and lower bounds, respectively. These latter uncertainties (under 95% credible intervals) are factored in with the second error bar (dotted lines in panel b of Fig. 5).

5. Discussion

5.1. Policy implications

The integration of spatially explicit household survey information and a method for predicting livestock distribution is essential for improving agricultural productivity and sustainability in Malawi and the wider African context. Accurate livestock location data is critical for regional policy makers. It enables them to design and implement targeted policies, such as improving the scope and reach of extension services, preventive disease control measures and environmental management strategies. For example, the dissemination of information on best practices, the orchestration of timely vaccination programmes and the management of grazing patterns can be tailored to the geographical distribution of livestock. This is supported by recent research, such as Rahimi et al. (2021), which emphasizes the influence of climate change on livestock and the consequent need for regional policy recalibration.

In addition, spatially detailed livestock maps are a critical component of risk assessments for natural hazards, particularly floods and heat events. By overlaying livestock distribution data with climate risk models, high-risk areas can be pinpointed with greater accuracy. For example, our application shows that livestock in Salima and Karonga districts are vulnerable to flooding — a concern echoed by global projections that suggest a significant proportion of the world's livestock could experience increased heat stress due to climate change (Thornton et al., 2021).

5.2. Strengths and weaknesses

The methodology used to produce up-to-date maps based on household surveys represents a major strength in terms of data accuracy and accessibility. The reliance on these surveys, which are increasingly available but not ubiquitously geocoded, represents a significant advance over the Global Livestock of the World (GLW) approach. GLW often relies on sub-national data that are not consistently available across countries, limiting its usefulness for single-country studies. In addition, GLW is a global product, of coarser resolution, and is necessarily published with a time lag due to the effort required to collect reliable data from around the world. In contrast, the approach presented here can be easily implemented on a country-by-country basis as soon as household survey results are available. In addition, the use of Bayesian statistics enhances the model by providing intervals for uncertainty analysis. These are a direct by-product of the methods employed and, as demonstrated in our application, directly improve risk assessment by providing a more transparent picture of the uncertainties involved in statistical modeling.

Conversely, the weaknesses of this approach are mainly due to data limitations. One notable issue is the intentional offsetting of geocodes to protect the confidentiality of farmers' locations. While this is necessary for privacy reasons, it can introduce bias, especially when integrating additional predictors such as land cover from satellite imagery or tsetse species distribution — variables with significant explanatory power when linked to precise geolocations. Another limitation is the poor extrapolation capability of the spatial predictions beyond national boundaries, which requires comprehensive coverage of the sampling area to ensure model reliability. Such limitations need to be carefully acknowledged and mitigated where possible to enhance the credibility of livestock distribution predictions.

5.3. Future research

The scalability of our mapping approach to other countries depends on the availability of a substantial number of geo-coded observations. The LSMS-ISA survey, a comprehensive household dataset covering approximately 768 enumeration areas, provides a robust basis for the current study. However, the applicability of our methods to smaller surveys, which inherently contain fewer observations, remains uncertain. A viable avenue for future research could be the merging of different household survey datasets (as in Blumenstock et al. (2015)). This strategy could compensate for limited data points and increase the applicability of the model in different geographical contexts. In addition, exploring the synergy between subnational data, similar to that used in the Global Livestock of the World (GLW) study, and point data could potentially refine the spatial prediction process, resulting in more accurate livestock distribution maps.

Future research could also extend to comprehensive climate risk assessments. Drawing parallels with our application to flood risk, the integration of heat stress maps with livestock distribution data could facilitate the quantification of livestock at risk from heat-related climate extremes. Following the approach of studies on heat stress and labor productivity (Parsons et al., 2021), there is scope to investigate the impact of climate change on livestock productivity. Such research would not only contribute to climate change adaptation, but also inform agricultural policies to ensure the resilience of livestock production systems to climate variability.

6. Conclusions

This paper presents a methodological framework for mapping the distribution of livestock in data-poor countries. The method is based on geo-located data from the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) survey and Bayesian spatial statistical modeling. We demonstrate the usefulness of our approach by estimating high-resolution 1 km^2 maps of livestock distribution for Malawi in 2010 and 2019, complemented by a set of spatial predictors.

Our results illustrate significant shifts in livestock distribution within Malawi over the study period, with particular growth in the small ruminant sector and variations in poultry distribution. We also compare the resulting livestock distributions with the Gridded Livestock of the World (GLW) database. The comparison with the survey site highlights the importance of using localized data sources to improve accuracy, as the GLW often underpredicts in critical locations. To demonstrate the policy relevance of our method, we identify where livestock populations are at risk from climate-induced flood events, highlighting the practical implications of this research for risk assessment and how it can assist in the formulation of effective agricultural policies.

However, it is worth noting that this study focusses on a linear predictor for the projection exercise. An interesting avenue for future research is the use of more nonlinear approaches. These methods could offer greater flexibility in capturing the complex relationships between spatial predictors and livestock distributions.

CRediT authorship contribution statement

Tamás Krisztin: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Michiel van Dijk:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Philipp Piribauer:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

The research carried out in this paper was supported by funds of the Oesterreichische Nationalbank (Jubilaeumsfond project number 18307). All authors reviewed and approved the final version of the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.envdev.2025.101141.

Data availability

Data will be made available on request.

References

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Bayesian Models with R-INLA, vol. 7, John Wiley & Sons, Ltd., Chichester, UK.

Blumenstock, J., Cadamuro, G., On, R., 2015. Predicting poverty and wealth from mobile phone metadata. Science 350 (6264), 1073–1076.

Cecchi, G., Paone, M., Argilés Herrero, R., Vreysen, M., Mattioli, R.C., 2015. Developing a continental atlas of the distribution and trypanosomal infection of tsetse flies (glossina species). Parasites & Vectors 8, 1–10.

Chiriacò, M.V., Valentini, R., 2020. A land-based approach for climate change mitigation in the livestock sector. J. Clean. Prod..

Cressie, N., Wikle, C.K., 2015. Statistics for Spatio-Temporal Data. John Wiley & Sons.

Defourny, P. abd G. Kirches, Brockmann, C., Boettcher, M., Peters, M., Bontemps, S., Lamarche, C., Schlerf, M., Santoro, M., 2012. Land cover cci. Prod. User Guid. Version 2 (325), 10–1016.

Eisen, M.B., Brown, P.O., 2021. Rapid global phaseout of animal agriculture has the potential to stabilize greenhouse gas levels for 30 years and offset 68 percent of co2 emissions. bioRxiv.

Ekwem, D., Morrison, T.A., Reeve, R., Enright, J., Buza, J., Shirima, G., Mwajombe, J.K., Lembo, T., Hopcraft, J.G.C., 2021. Livestock movement informs the risk of disease spread in traditional production systems in east africa. Sci. Rep. 11 (1), 16375.

Food and Agriculture Organization of the United Nations, 2022. Livestock sector report - a review of the livestock sector in malawi in 2021. http://dx.doi.org/10.4060/cc1073en.

Gilbert, M., Conchedda, G., Van Boeckel, T.P., Cinardi, G., Linard, C., Nicolas, G., Thanapongtharm, W., D'Aietti, L., Wint, W., Newman, S.H., Robinson, T.P., 2015. Income disparities and the global distribution of intensively farmed chicken and pigs. PLOS ONE 10 (7), e0133381.

Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T.P., Vanwambeke, S.O., Wint, G., Robinson, T.P., 2018. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. Sci. Data 5 (1), 1–11.

Godde, C.M., Mason-DCroz, D., Mayberry, D., Thornton, P.K., Herrero, M., 2021. Impacts of climate change on the livestock food supply chain; a review of the evidence. Glob. Food Secur. 28, 100488.

IUCN, U.-W., et al., 2010. The world database on protected areas (wdpa). Annual release. pp. 1–9.

Julien, J.C., Bravo-Ureta, B., Rada, N.E., 2019. Assessing farm performance by size in malawi, tanzania, and uganda. Food Policy.

Kaumbata, W., Banda, L., Mészáros, G., Gondwe, T., Woodward-Greene, M., Rosen, B.D., Van Tassell, C.P., Sölkner, J., Wurzinger, M., 2020. Tangible and intangible benefits of local goats rearing in smallholder farms in malawi. Small Ruminant Res. 187, 106095.

Krisztin, T., Piribauer, P., 2023. A bayesian approach for the estimation of weight matrices in spatial autoregressive models. Spat. Econ. Anal. 18 (1), 44–63.

Krisztin, T., Piribauer, P., Wögerer, M., 2022. A spatial multinomial logit model for analysing urban expansion. Spat. Econ. Anal. 17 (2), 223–244.

Lam, C., Souza, P.C., 2020. Estimation and selection of spatial weight matrix in a spatial lag model. J. Bus. Econom. Statist. 38 (3), 693–710.

Laurini, M.P., 2017. A spatial error model with continuous random effects and an application to growth convergence. J. Geogr. Syst. 19 (4), 371–398.

LeSage, J., Pace, R.K., 2009. Introduction To Spatial Econometrics. Chapman and Hall/CRC..

Lindgren, F., Rue, H., 2015. Bayesian Spatial Modelling with R-INLA. J. Stat. Softw. 63 (19), 1–26.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (4), 423–498.

Meijer, J.R., Huijbregts, M.A., Schotten, K.C., Schipper, A.M., 2018. Global patterns of current and future road infrastructure. Environ. Res. Lett. 13 (6), 064006.

Moraga, P., Dean, C., Inoue, J., Morawiecki, P., Noureen, S.R., Wang, F., 2021. Bayesian spatial modelling of geostatistical data using inla and spde methods: A case study predicting malaria risk in mozambique. Spat. Spatio- Temporal Epidemiol. 39, 100440.

NASA JPL, 2013. NASA shuttle radar topography mission global 1 arc second. http://dx.doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003, (Accessed 03 December 2020) from NASA EOSDIS Land Processes DAAC.

Neumann, K., Elbersen, B.S., Verburg, P.H., Staritsky, I., Pérez-Soba, M., de Vries, W., Rienks, W.A., 2009. Modelling the spatial distribution of livestock in Europe. Landscape Ecol. 24 (9), 1207–1222.

Ng'ong'ola, D.H., Maonga, B.B., Makoka, D., Chirwa, C.W., 2017. Impact of malawi's farm income diversification programme on household welfare: Empirical evidence from eleven districts. Food Policy Res. Inst. 20.

Nicolas, G., Robinson, T., Wint, W., Conchedda, G., Cinardi, G., Gilbert, M., 2016. Using random forest to improve the downscaling of global livestock census data. PLoS ONE 11.

North, M.A., Franke, J.A., Ouweneel, B., Trisos, C.H., 2023. Global risk of heat stress to cattle from climate change. Environ. Res. Lett. 18 (9), 094027.

Palmer, P.I., Wainwright, C.M., Dong, B., Maidment, R.I., Wheeler, K.G., Gedney, N., Hickman, J.E., Madani, N., Folwell, S.S., Abdo, G., et al., 2023. Drivers and impacts of eastern african rainfall variability. Nat. Rev. Earth Environ. 4 (4), 254–270.

Parsons, L.A., Shindell, D., Tigchelaar, M., Zhang, Y., Spector, J.T., 2021. Increased labor losses and decreased adaptation potential in a warmer world. Nat. Commun. 12 (1), 7286.

Piribauer, P., Glocker, C., Krisztin, T., 2023. Beyond distance: The spatial relationships of European regional economic growth. J. Econom. Dynam. Control 155, 104735.

Prosser, D.J., Wu, J., Ellis, E.C., Gale, F., Van Boeckel, T.P., Wint, W., Robinson, T.P., Xiao, X., Gilbert, M., 2011. Modelling the distribution of chickens, ducks, and geese in China. Agricult. Ecosys. Environ. 141 (3–4), 381–389.

Rahimi, J., Mutua, J., Notenbaert, A., Marshall, K., ButterbachBahl, K., 2021. Heat stress will detrimentally impact future livestock production in east africa. Nat. Food 2, 88–96.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., Rafaj, P., 2011. Rcp 8.5—a scenario of comparatively high greenhouse gas emissions. Clim. Change 109, 33–57.

Robinson, T.P., Thornton, P.K., Franceschini, G., Kruska, R., Chiozza, F., Notenbaert, A., Cecchi, G., Herrero, M., Epprecht, M., Fritz, S., You, L., Conchedda, G., See, L., 2011. Global Livestock Production Systems. Technical report, FAO, Rome.

Robinson, T.P., Wint, W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., Cinardi, G., D'Aietti, L., Hay, S.I., Gilbert, M., 2014. Mapping the global distribution of livestock. PLoS ONE 9 (5), e96084.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman and Hall/CRC.

Rue, H., Martino, S., Chopin, N., 2009a. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392.

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

Santeramo, F.G., Lamonaca, E., Tappi, M., Gioia, L., 2020. On the environmental impacts of voluntary animal-based policies in the eu: Technical and political considerations. Environ. Sci. Policy.

Simpson, D., Lindgren, F., Rue, H., 2012. Think continuous: Markovian Gaussian models in spatial statistics. Spat. Stat. 1, 16–29.

Tatem, A.J., 2017. Worldpop, open data for spatial demography. Sci. Data 4 (1), 1–4.

Tellman, B., Sullivan, J.A., Kuhn, C., Kettner, A.J., Doyle, C.S., Brakenridge, G.R., Erickson, T.A., Slayback, D.A., 2021. Satellite imaging reveals increased proportion of population exposed to floods. Nature 596 (7870), 80–86.

Thornton, P., Nelson, G.C., Mayberry, D., Herrero, M., 2021. Increases in extreme heat stress in domesticated livestock species during the twenty-first century. Global Change Biol. 27, 5762–5772.

Uwizeye, A., de Boer, I.J.M., Opio, C., Schulte, R.P.O., Falcucci, A., Tempio, G., Teillard, F., Casu, F., Rulli, M.C., Galloway, J.N., Leip, A., Erisman, J.W., Robinson, T.P., Steinfeld, H., Gerber, P.J., 2020. Nitrogen emissions along global livestock supply chains. Nat. Food 1, 437–446.

Van Boeckel, T.P., Prosser, D.J., Franceschini, G., Biradar, C., Wint, W., Robinson, T.P., Gilbert, M., 2011. Modelling the distribution of domestic ducks in monsoon Asia. Agricult. Ecosys. Environ. 141 (3–4), 373–380.

Van Boeckel, T.P., Thanapongtharm, W., Robinson, T.P., D'Aietti, L., Gilbert, M., 2012. Predicting the distribution of intensive poultry farming in thailand. Agricult. Ecosys. Environ. 149, 144–153.

Van Dijk, M., Morley, T., van Loon, M., Reidsma, P., Tesfaye, K., van Ittersum, M.K., 2020. Reducing the maize yield gap in ethiopia: Decomposition and policy simulation. Agricult. Sys. 183, 102828.

Ward, P.J., Winsemius, H.C., S. Kuzma, M.F. Bierkens, Bouwman, A., De Moel, H., Loaiza, A.D., Eilander, D., Englhardt, J., Erkens, G., et al., 2020. Aqueduct Floods Methodology. World Resources Institute, pp. 1–28.

Weiss, D.J., Nelson, A., Gibson, H., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., et al., 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature 553 (7688), 333–336.

Wint, W., Robinson, T.P., 2007. Gridded Livestock of the World. Technical report, Rome.

World Bank Group, 2022. Country Climate and Development Report: Malawi. Technical Report. 202-473-1000; Internet: www.worldbank.org, World Bank Group, 1818 H Street NW, Washington DC 20433. Telephone.

Zomer, R.J., Xu, J., Trabucco, A., 2022. Version 3 of the global aridity index and potential evapotranspiration database. Sci. Data 9 (1), 409.