




OPEN

## A time-series approach to mapping livestock density using household survey data

Julianne Meisner<sup>1,2</sup>, Agapitus Kato<sup>3,11</sup>, Marshall Lemerani<sup>4,11</sup>, Erick Mwamba Miaka<sup>5,11</sup>, Acaga Taban Ismail<sup>6,11</sup>, Jonathan Wakefield<sup>7,8</sup>, Ali Rowhani-Rahbar<sup>2</sup>, David Pigott<sup>9,10</sup>, Jonathan Mayer<sup>2</sup> & Peter Rabinowitz<sup>1,10</sup>

More than one billion people rely on livestock for income, nutrition, and social cohesion, however livestock keeping can facilitate disease transmission and contribute to climate change. While data on the distribution of livestock have broad utility across a range of applications, efforts to map the distribution of livestock on a large scale are limited to the Gridded Livestock of the World (GLW) project. We present a complementary effort to map the distribution of cattle and pigs in Malawi, Uganda, Democratic Republic of Congo, and South Sudan. In contrast to GLW, which uses dasymetric modeling applied to census data to produce time-stratified estimates of livestock counts and spatial density, our work uses complex survey data and distinct modeling methods to generate a time-series of livestock distribution, defining livestock density as the ratio of animals to humans. In addition to favorable cross-validation results and general agreement with national density estimates derived from external data on national human and livestock populations, our results demonstrate extremely good agreement with GLW-3 estimates, supporting the validity of both efforts. Our results furthermore offer a high-resolution time series result and employ a definition of density which is particularly well-suited to the study of livestock-origin zoonoses.

Globally, one billion people living on less than US\$2 per day depend on livestock. For these communities, which include 80% of the poor in Africa, livestock are a critical source of household income, transport, draft power for crop agriculture, and nutrition, providing 11% of energy and 26% of dietary protein among the poor in East Africa, and up to 50% of energy for children under five in pastoralist communities<sup>1,2</sup>. Livestock keeping also plays an important cultural role, serving as a means to maintain family cohesion and social networks, gain political prestige, and strengthen legal claims on pasture land<sup>3,4</sup>. These critical roles livestock play, however, commonly result in close human-animal contact, driving transmission of zoonotic diseases and resulting in over 2.5 billion cases of human illness and 2.7 million deaths per year<sup>2</sup>.

Livestock also have major environmental impacts across a range of scales. In Africa, where an estimated 1 billion or more of the projected increase in the human population will occur, urbanization and an increased demand for animal products are also expected<sup>5</sup>, driving the so-called “livestock revolution<sup>6</sup>.” Production systems are generally low-input in sub-Saharan Africa, with very little technological change in the past 40 years<sup>5</sup>; compounded by degraded natural resources, this lack of technological change requires increasing production demands to be met by overgrazing and land-use changes, rather than intensification, compromising biodiversity and ecosystem services<sup>7</sup>. Furthermore, where intensification has been achieved, animals and animal wastes are generally concentrated, facilitating disease transmission and water pollution in the absence of proper controls<sup>1</sup>.

On a larger scale, livestock systems have a bi-directional relationship with climate change. Climate change affects the quantity and quality of feeds available to livestock, drives production losses due to heat stress, promotes

<sup>1</sup>Department of Environmental and Occupational Health Sciences, Center for One Health Research, University of Washington, Seattle 98195, USA. <sup>2</sup>Department of Epidemiology, University of Washington, Seattle 98195, USA. <sup>3</sup>Uganda Virus Research Institute, Entebbe, Uganda. <sup>4</sup>Ministry of Health, Lilongwe, Malawi. <sup>5</sup>Programme National de Lutte contre la Trypanosomiase Humaine Africaine, Kinshasa, Democratic Republic of Congo. <sup>6</sup>IntraHealth International, Juba, South Sudan. <sup>7</sup>Department of Biostatistics, University of Washington, Seattle 98195, USA. <sup>8</sup>Department of Statistics, University of Washington, Seattle 98195, USA. <sup>9</sup>Department of Health Metrics Sciences, University of Washington, Seattle 98195, USA. <sup>10</sup>Department of Global Health, University of Washington, Seattle 98195, USA. <sup>11</sup>These authors contributed equally: Agapitus Kato, Marshall Lemerani, Erick Mwamba Miaka and Acaga Ismail Taban. ✉email: meisnerj@uw.edu

uneven distribution of water resources, and modifies the distribution of livestock diseases and disease vectors, representing another manifestation of the disproportionate burden climate change places on the world's resource poor. In turn, livestock and livestock systems are substantial users of natural resources, in particular water and land for grazing and feed production, and contribute to greenhouse gas emissions and climate change<sup>5,8</sup>.

Thus, high spatiotemporal resolution maps of the distribution of livestock hold utility for a range of research and policy applications, from epidemiology and public health to economics and climate science. To this end, the Gridded Livestock of the World (GLW) database was developed in 2007<sup>9</sup>, with GLW-3 being the most recent iteration. Published in 2018 but providing results for 2010, GLW-3 produces two sets of results, one using dasymetric modeling to disaggregate census counts, which is based on weights derived from statistical models that use spatial covariates, and a simple aerial weighting approach that produces estimates free from association with the spatial covariates used in the dasymetric results. These results are produced globally at a resolution of 0.083333° for counts of cattle, buffaloes, horses, sheep, goats, pigs, chickens, and ducks<sup>10</sup>. GLW does not use any household survey data, such as the Demographic and Health Survey or the Living Standards Measurement Studies, as these data use complex survey sampling approaches that would not be well-suited to GLW's current modeling approaches. While numerous prior authors have mapped livestock diseases and disease vectors, beyond GLW mapping of livestock itself remain isolated to highly localized efforts such as grazing intensity in Kazakhstan<sup>11</sup> and livestock movements in Sahelian Africa<sup>12</sup>.

To build on and support GLW efforts, we have used household survey and census data on livestock ownership and the stochastic partial differential equations (SPDE) approach to Gaussian process modeling to generate high-resolution (0.017°) maps of cattle and pig density. In contrast to the dasymetric modeling used by GLW, this approach can be applied to both household survey data and census data, expanding the data available for modeling. This approach also leverages spatial structure to strengthen predictions, and generates principled uncertainty estimates derived from posterior distributions. We have produced these maps for every year from 2000–2020 for Malawi and Uganda and 2008–2015 for Democratic Republic of Congo (DRC), facilitating their use in longitudinal analyses. We have additionally used small area estimation to produce density maps at the county level for South Sudan in 2008. In all four countries, we define density as the ratio of animals to humans, producing a map in which the human population is flattened. We present here our methods and resulting products, and discuss our results in the context of GLW-3 and extension of these approaches to other countries and species.

## Results

**Input data.** In the final dataset there were 6330 clusters for cattle mapping and 6342 clusters for pig mapping in Malawi; 3269 clusters for cattle and 3310 clusters for pigs in Uganda; and 861 clusters for each species in DRC. For South Sudan, data were available for all counties except for Khorflus. For the three SPDE countries (Malawi, Uganda, and DRC), cluster distribution in space is presented in Fig. 1, and data richness (number of survey clusters) by year and country is summarized in Fig. 2.

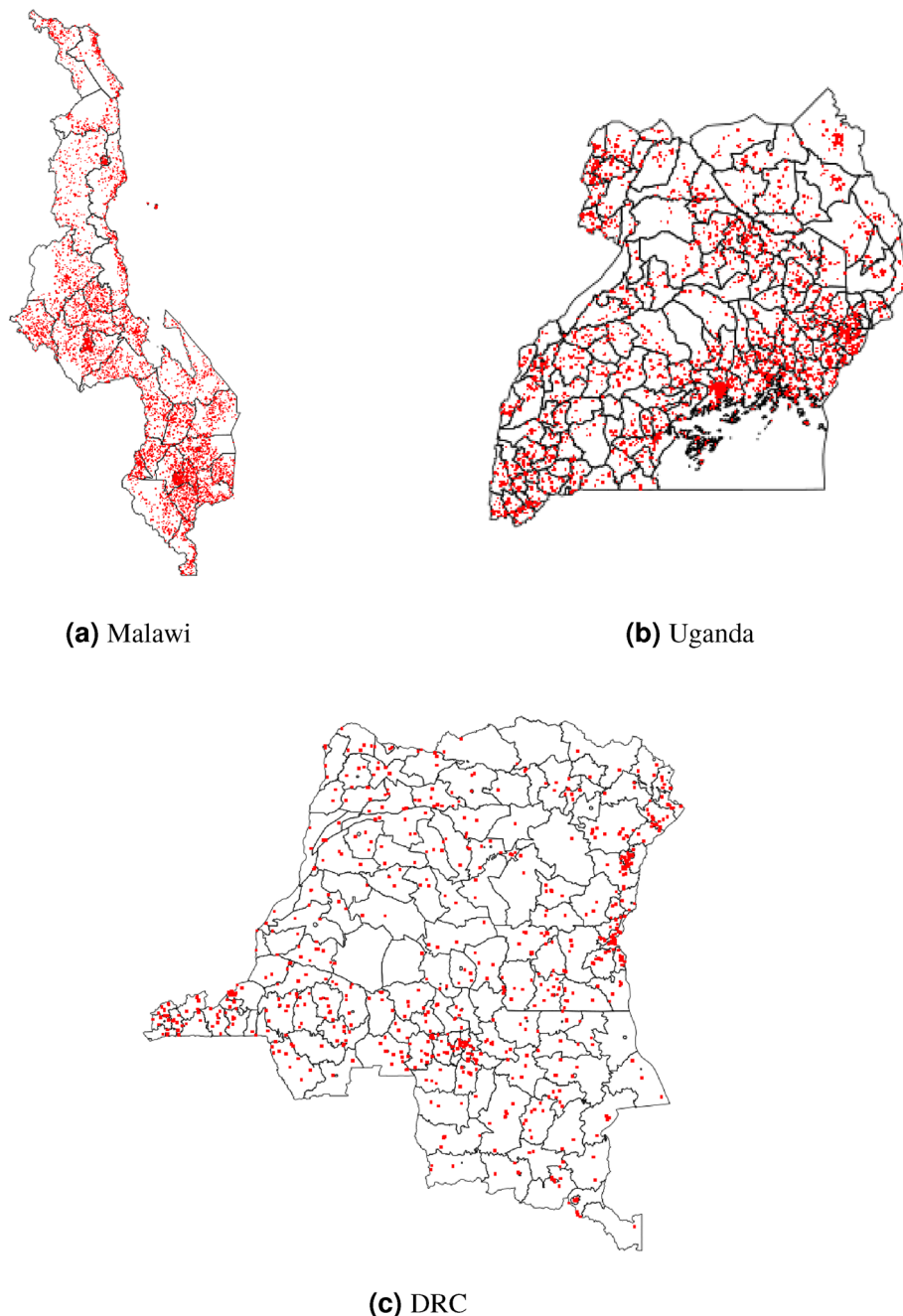
**Model selection (SPDE only).** Leave-one-out cross-validation (LOOCV) results are presented in Fig. 3; we have removed models and surveys with very high mean squared error (MSE) for the sake of interpretability of these figures—as detailed in the accompanying caption—however these surveys and models were included when examining overall model performance. Candidate models differed in fixed effects—covariates—and flexibility of the temporal random effect. For Uganda the best-fitting model was model 6 (MSE 0.076) which included all candidate covariates (urban/rural status, protected areas, elevation, and bodies of water) and a less flexible temporal random effect (random walk 1). For Malawi, model 3 was the best fit (MSE 0.023) which included urban/rural status, protected areas, and a more flexible temporal random effect (random walk 2), and for DRC model 2 (urban/rural status only, random walk 2) was the best fit (MSE 0.048).

In all three countries, we found extremely strong positive associations between our livestock density estimates and those derived from GLW-3 (coefficients greater than 100), for both cattle and pig density, after using WorldPop<sup>13</sup> data as a denominator to convert livestock counts estimated by GLW-3 to densities. These remarkably high coefficients can be explained by the coarser resolution of GLW-3 (0.083°) compared with both our maps (0.017°) and WorldPop data (0.0083°), resulting in markedly higher density estimates for GLW-3. They do, however, indicate general agreement between our estimates and those of GLW-3.

**Random effects posteriors.** For the marginal standard deviation, the selected prior yields a posterior 95% credible interval of (0.37, 2.72) for the residual rate ratio (deviations in livestock density above or below the mean model). For the iid random effects, the specified priors result in a posterior 99% credible interval for the residual variance of (0.6, 1.6).

**Maps.** Density maps for 2005 and 2010 for Malawi and Uganda, 2010 and 2015 for DRC, and for 2008 for South Sudan, are presented in Figs. 4, 5, 6, 7, 8, 9, 16. Maps for 2010 are presented next to corresponding GLW-3 products in Figs. 10, 11, 12, 13, 14, 15, and 17, 18. Results (median and width of the posterior 95% credible interval) as .tif files for every year for Malawi, Uganda, and DRC, and as shapefiles for 2008 for South Sudan, are available in a GitHub repository: <https://github.com/JulianneMeisnerUW/LivestockMaps>. Providing estimates of both median and uncertainty allows users to assess the quality of the estimates for each species, country, and year. All R code used to generate these maps are also in this repository.

In Malawi (Figs. 4, 5), cattle density was highest in the northern and southern extents of the country, and pig density generally decreased along a west-east gradient. Across all years and clusters, median cattle density was 0.05, and median pig density was 0.07. Per FAOSTAT total stock estimates<sup>14</sup> and World Bank national population

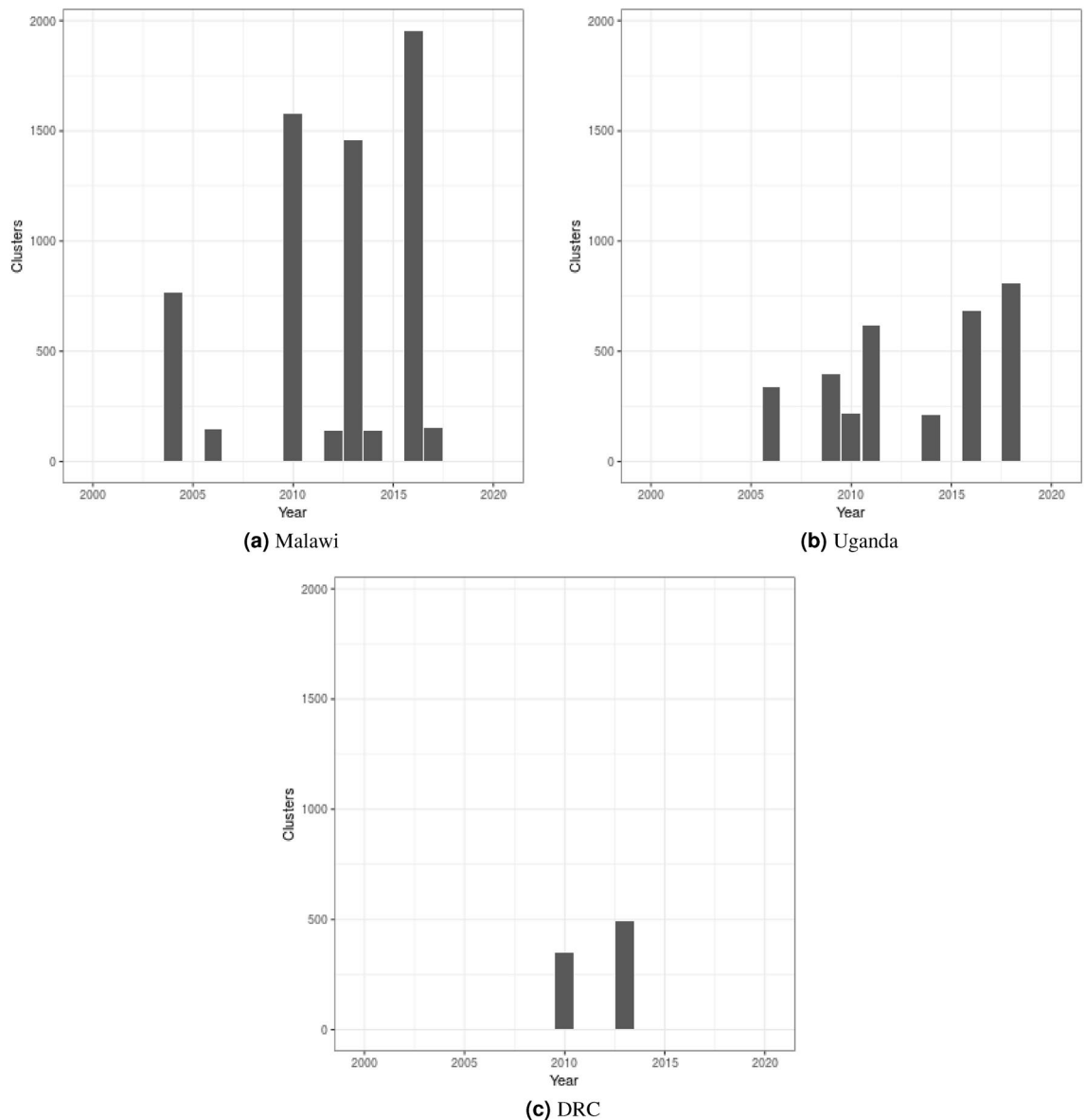


**Figure 1.** Spatial distribution of survey clusters (observed data) across all years. Size of the red dots does not scale with data richness.

estimates<sup>15</sup>, mean national cattle density over the period 2000–2019 was 0.08, and mean national pig density was 0.15. Spatial patterns appear to largely agree across our maps and GLW-3 maps cropped to Malawi (Figs. 10, 11); exact agreement is not expected as these maps are parameterized differently.

In Uganda (Figs. 6, 7), the “cattle corridor” is clearly visible in both years, while pig density appears to be higher in the south of Uganda. Across all years and clusters, median cattle density was 0.23, and median pig density was 0.02. Per FAOSTAT total stock estimates<sup>14</sup> and World Bank national population estimates<sup>15</sup>, mean national cattle density over the period 2000–2019 was 0.33, and mean national pig density was 0.07. Spatial patterns are similar to those of GLW-3 cropped to Uganda (Figs. 12, 13), however two GLW-3 hotspots—a cattle density hotspot in the northeast and a pig density hotspot in the eastern part of the northern region—were not found in our maps, and pig density hotspots in the Central Region detected in our maps was not detected in GLW-3; again, we do not expect perfect agreement between these products.

In DRC, there are “hotspots” of cattle density in the east and southwest of the country, and no obvious spatial patterns for pig density (Figs. 8, 9) Per FAOSTAT total stock estimates<sup>14</sup> and World Bank national population estimates<sup>15</sup>, mean national cattle density over the period 2008–2015 was 0.013 for cattle and 0.014 for pigs, while



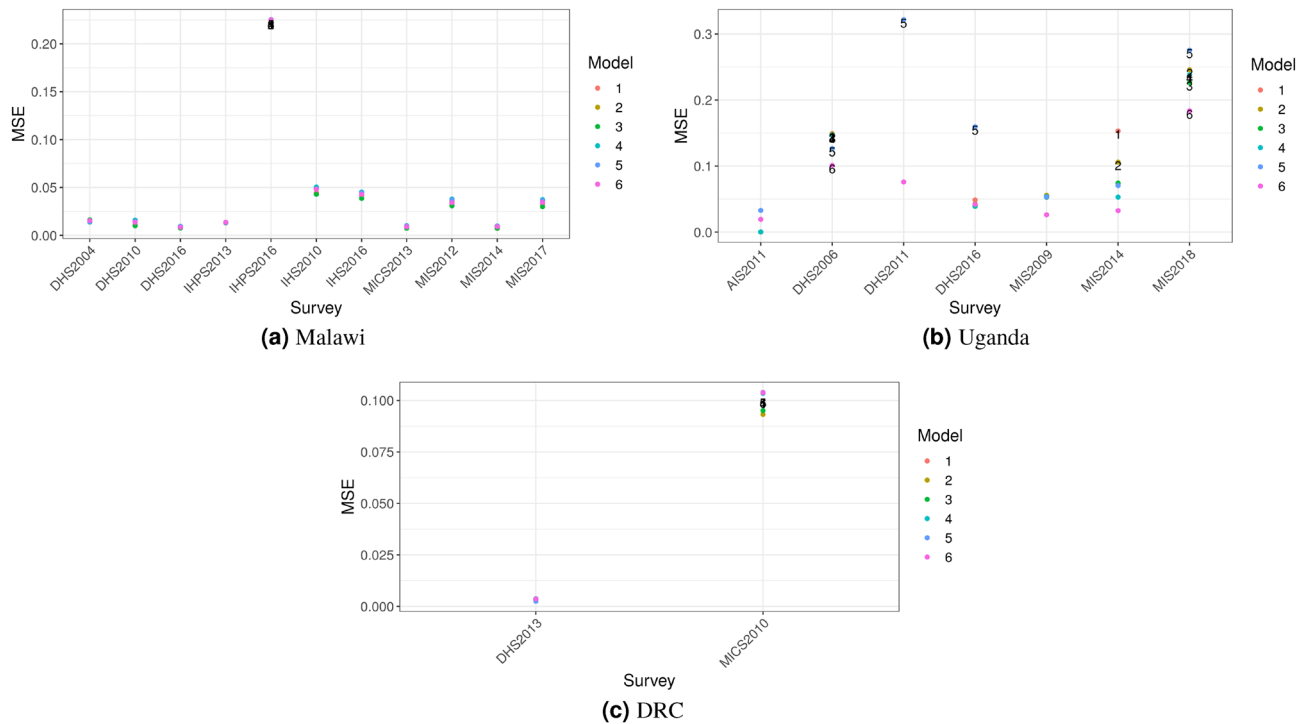
**Figure 2.** Number of survey clusters (enumeration areas) available for cattle density mapping in each year and country. Results were nearly equivalent for pigs.

across all years and clusters our median estimate was 0.037 for cattle and 0.053 for pigs. Spatial patterns are difficult to detect in the GLW-3 cattle product for DRC (Fig. 14), however a density hotspot in the northeast is evident in both maps. In the pig maps, a density hotspot in west-central DRC that is evident in both products (Fig. 15).

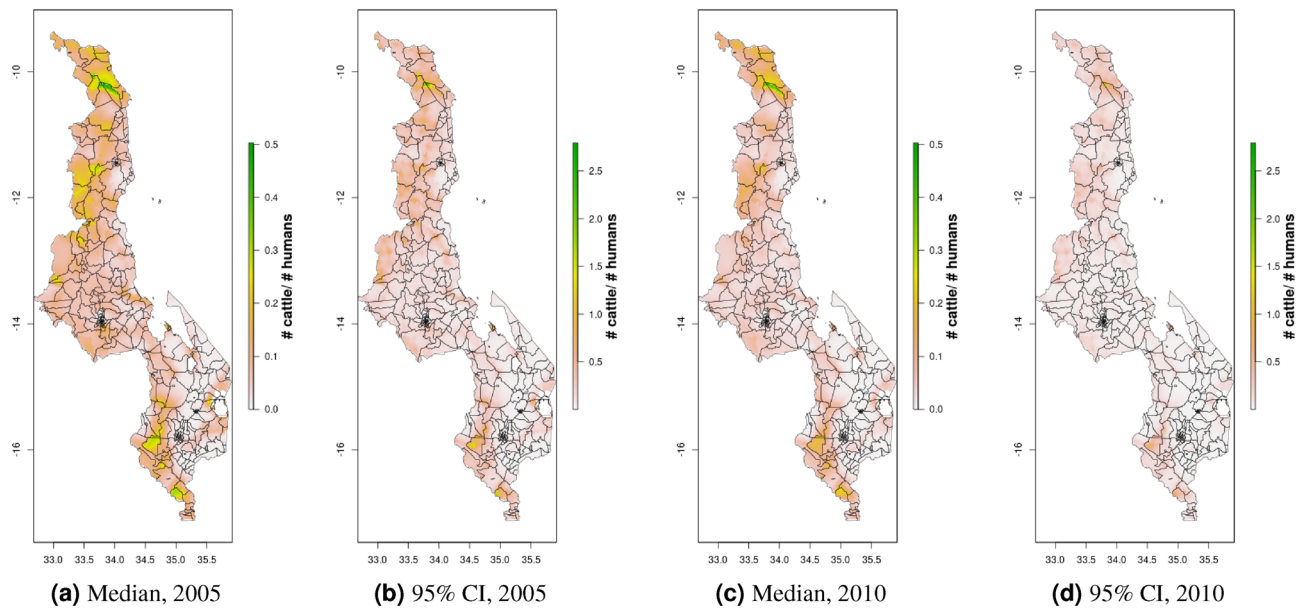
Notably, in Malawi, DRC, and Uganda, a general decrease in cattle density is observed over time across the two plotted years (2005 and 2010 for Uganda and Malawi, 2010 and 2015 for DRC). For pigs this is observed to a much lesser extent in Malawi, and not at all in Uganda and DRC. This likely represents an increase in human population density that is matched by corresponding increases in pig density but not cattle density. Figures 10, 11, 12, 13, 14, 15 also present modeled and observed density for 2010 in each country. Agreement appears good overall, with the apparent exception of cattle in Uganda (Fig. 12), where several observed foci of high density are not captured in the modeled surface. This can likely be attributed to observations from adjacent years (2009 and 2011), which will influence 2010 estimates through the temporal random effect, and are not represented in this time stratified figure.

In South Sudan (Fig. 16), cattle density was on average higher than pig density, and there is a clear decreasing southeast-northwest trend in cattle density. For pig density in South Sudan, other than generally low densities in the northwest of the country, there are no clear spatial trends. Across all counties, median density was 1.56 for cattle and 0.007 for pigs in 2008. Per FAOSTAT total stock estimates<sup>14</sup> and World Bank national population estimates<sup>15</sup>, mean national cattle density over the period 2000–2019 was 1.33; no FAOSTAT entry is available for pigs in South Sudan. On comparison with GLW-3 maps cropped to South Sudan, “hotspots” of cattle density are more to the east in our maps than GLW-3 maps (Fig. 17), and pig density is markedly different in our maps from count distribution in GLW-3 (Fig. 18).





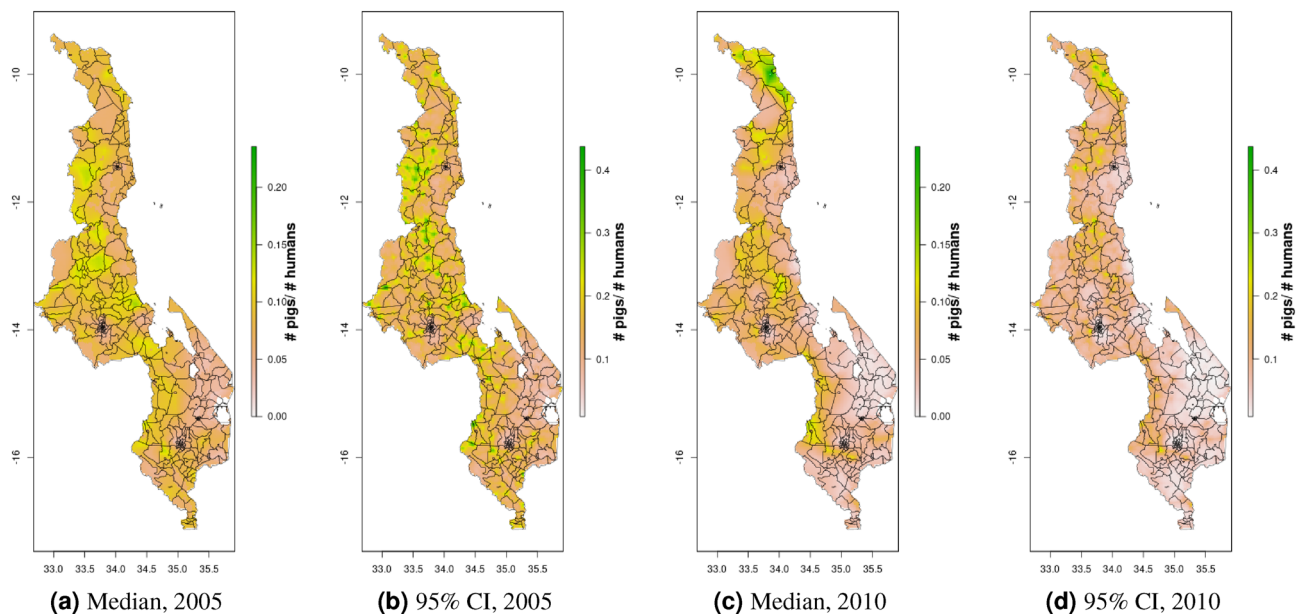
**Figure 3.** Leave one out cross validation results. Models with mean squared error (MSE) > 0.1 are labeled. Removed from figures (but not MSE calculations) for interpretability: (a) Malawi: 2004 Integrated Household Survey, all models; Uganda (b): 2009 National Panel Survey (NPS), all models, 2011 DHS, models 1-2, 2010 NPS, models 1 and 3-6, and 2011 NPS, models 1-3 and 5.



**Figure 4.** Cattle density mapping results, Malawi, 2005 and 2010. Median (a, c) and width of posterior 95% credible interval (b, d).

### Discussion

We have generated a set of cattle and pig density maps for Malawi, DRC, Uganda, and South Sudan, publicly available for use in economic, public health, environmental, agricultural, and other applications. Our results compliment GLW-3 efforts by producing a longitudinal product in Malawi, DRC, and Uganda, parameterizing livestock as the ratio of heads of cattle and pigs to the human population, and utilizing distinct input data and an entirely different modeling approach.



**Figure 5.** Pig density mapping results, Malawi, 2005 and 2010. Median (a, c) and width of posterior 95% credible interval (b, d).

In contrast with GLW-3, we did not mask any pixels as unsuitable for livestock. The result is our findings impose no assumptions with regards to livestock presence in urban and high-altitude pixels, livestock incursion into protected areas, and wet versus dry season changes in the distribution of water bodies, but may result in bias in pixels corresponding to large and permanent bodies of water (Fig. 20).

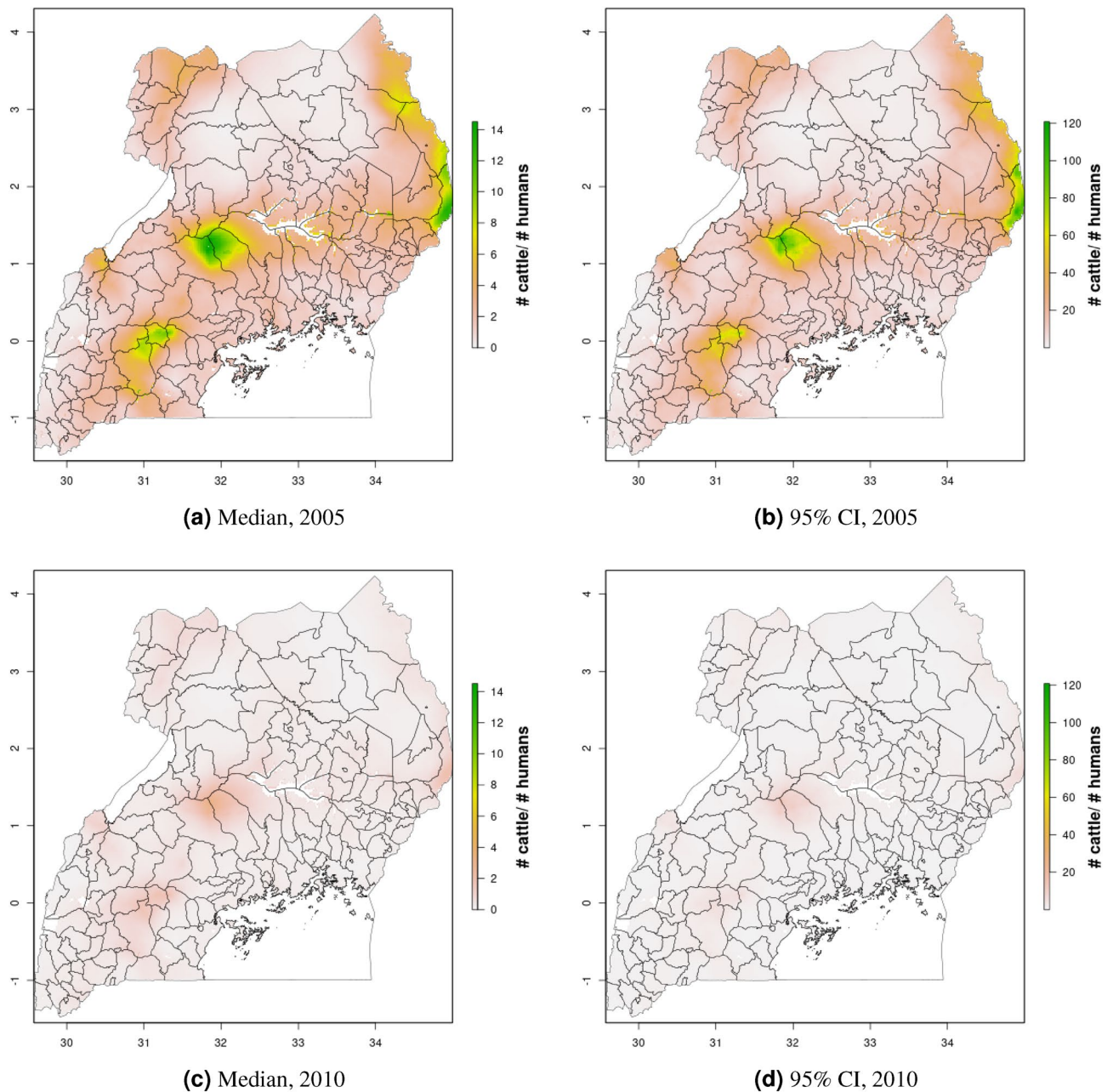
There are several limitations to our results. First, there may be inaccuracies in the input data, for instance over-reporting of livestock ownership due to the social desirability of owning a larger herd, or under-reporting due to concerns about increased taxation. Other inaccuracies may arise due to jittering in the input data, in which clusters in publicly-available datasets are randomly displaced to protect participant privacy. In light of the high resolution of our product (approx. 2 km<sup>2</sup> at the equator), this is particularly problematic in rural areas, where displacement is up to 5 km in DHS data, and a random 1% of clusters are displaced up to 10 km<sup>17</sup>. To our knowledge methods for probabilistically “un-jittering” data are under development but not yet available. While ground-truthing is both cost-prohibitive and impractical for a longitudinal product, our LOOCV results indicate low MSEs overall, and good agreement between our results and GLW-3 results is a promising indicator of the validity of both products. Furthermore, our findings generally agreed with national density estimates, and we would not expect perfect agreement with such estimates as local distribution patterns for humans is unlikely to be equivalent to local distribution patterns for livestock, and we are comparing medians to means. In addition to jittering in the input data sources, misclassification of livestock density may arise from the location definition used in our models, whereby livestock are assigned to the centroid of the corresponding enumeration area. Thus our maps should be interpreted as the density of livestock to humans when both are in their place of residence (i.e., at nighttime). While this may be a poor reflection of livestock presence in pastoralist systems, it is difficult to conceptualize any other definition of livestock density as the daily movements of humans and livestock may be distinct.

Second, our results treat livestock production systems as equivalent within a given species, however the environmental, socioeconomic, and public health relevance of production system types (e.g., pastoralist vs. intensive) are distinct, potentially reducing the utility of our results for select uses. Furthermore, users will need to take care to ensure the predictors (spatial covariates) used in our SPDE models do not introduce circularity in their analyses. We have opted for a parsimonious (few covariates) model and excluded vegetation indices as a predictor for this reason, however, users who wish to study the distribution of livestock with respect to, for instance, national parks, may find use of our maps biases their results, as boundaries of protected areas were used as a predictor in our model.

Third, our approach to generating urbanicity surfaces reflects the challenging nature of meaningfully assigning unsampled cluster-years as urban versus rural. Our approach performed well in the test data with the exception of urban clusters in Uganda, for which it performed very poorly. Alternative approaches, such as using Global Human Settlement Layer data<sup>18</sup> and other existing urbanicity surfaces, do not necessarily correspond to the definition of urbanicity needed in our application, nor are they generally available as a longitudinal product.

Fourth, due to limited data in DRC and South Sudan, we could not include a space-time interaction term in DRC—imposing the assumption that spatial trends are constant in time, and temporal trends are constant in space—and could only produce a county-level map for South Sudan, and only for a single year.

Finally, our estimates are accompanied by uncertainty which varies across time, countries, and clusters. Uncertainty was highest for DRC and for early years, rendering our maps less useful in these cluster-years. It



**Figure 6.** Cattle density mapping results, Uganda, 2005 and 2010. Median (a, c) and width of posterior 95% credible interval (b, d).

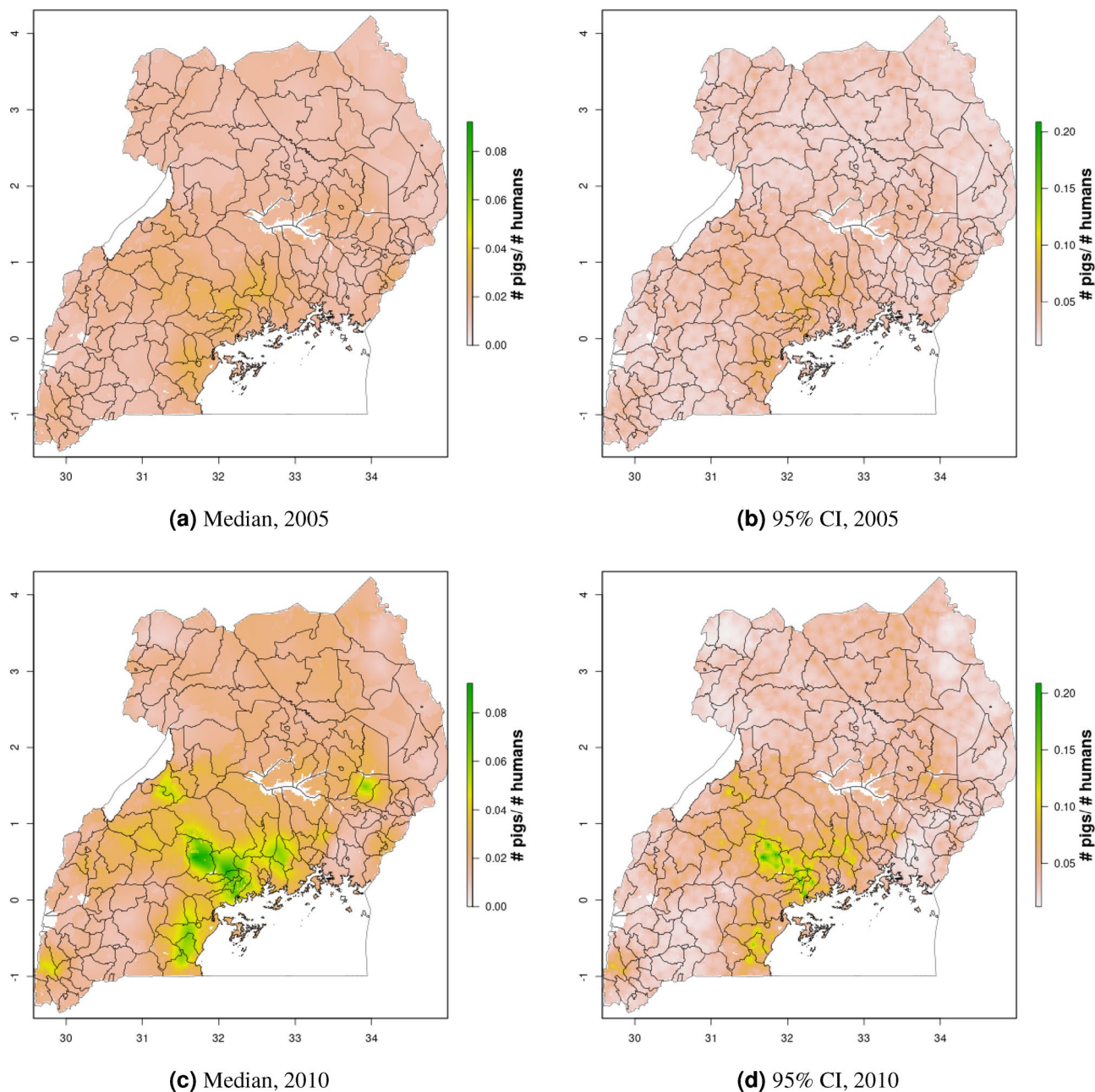
is our hope that by producing uncertainty estimates alongside of medians, it will be transparent to users where our results are more or less stable.

For poor livestock keepers throughout the world, livestock represent a lifeline, providing financial security, nutrition, transportation, and connection to cultural identities. Livestock, however, also represent a source of exposure to zoonoses, and a driver of environmental change on local and global scales. To grapple with these divergent effects in research and policy, high-resolution estimates of livestock distribution tethered to human population distribution are needed. We present such a product, which, despite its limitations, represents an important extension to GLW-3 with high demonstrated validity and broad potential utility.

## Methods

**Data collection and processing.** Data sources were identified by searching GHDx and the IHSN Central Data Catalog<sup>19,20</sup>, and microdata were downloaded from publicly-accessible websites or by request to relevant national agencies. Both Demographic and Health Survey (DHS) data and population and housing census data were downloaded from IPUMS<sup>21,22</sup>; most other surveys were downloaded from links in the relevant IHSN entry. Data were processed by source; data sources by country are detailed in Table 1.

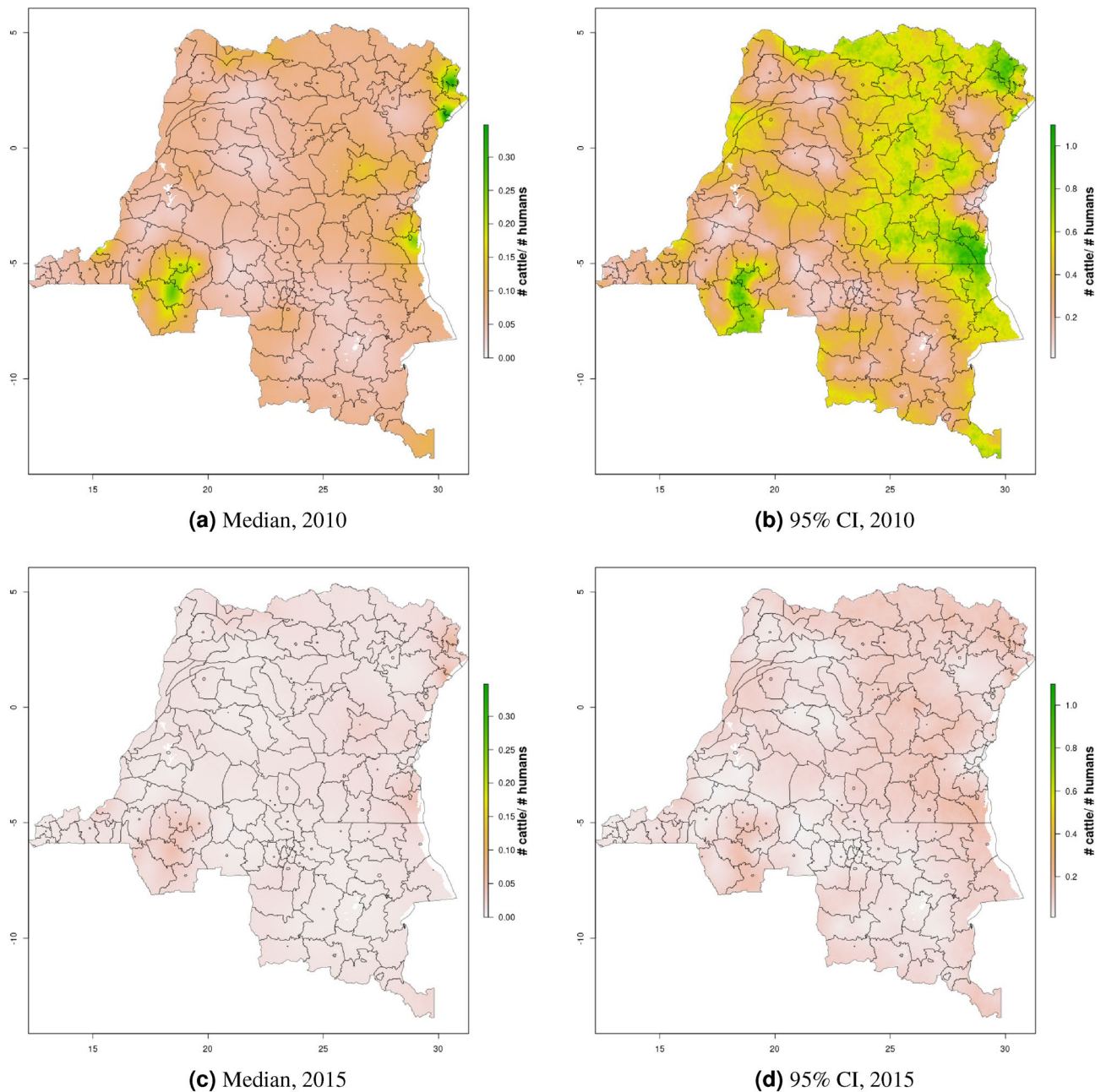




**Figure 7.** Pig density mapping results, Uganda, 2005 and 2010. Median (a, c) and width of posterior 95% credible interval (b, d).

As data were only available for a single year for South Sudan (2008), no longitudinal mapping could be performed for this country. Furthermore, these data provided geolocation only to the level of the county (administrative level 2), thus we have used small area estimation, rather than the SPDE approach, to generate county-level maps in South Sudan. In Malawi, Uganda, and DRC, survey cluster—the smallest geographical sampling unit of a household survey, typically an enumeration area consisting of approximately 100 households—was the unit of analysis.

Number of livestock was parameterized as number in sampled households, summed over cluster (Malawi, Uganda, DRC) or county (South Sudan). For cattle, counts were summed over local and exotic breeds, and dairy and non-dairy herds. Similarly, number of humans was parameterized as number of individuals (adults and children) in sampled households, summed over households in a given cluster (Malawi, Uganda, DRC) or county (South Sudan). In Malawi, DRC, and Uganda, clusters with missing values for cattle counts did not contribute to the denominator for cattle density, clusters with missing values for pig counts did not contribute to the denominator for pig density, and clusters missing both pig and cattle counts were removed. As surveys which used stratified sampling methods defined strata by geography and urban/rural status, urban/rural status was also extracted for later modeling. Year was defined as survey year; for surveys conducted over two years, the



**Figure 8.** Cattle density mapping results, DRC, 2010 and 2015. Median (**a, c**) and width of posterior 95% credible interval (**b, d**).

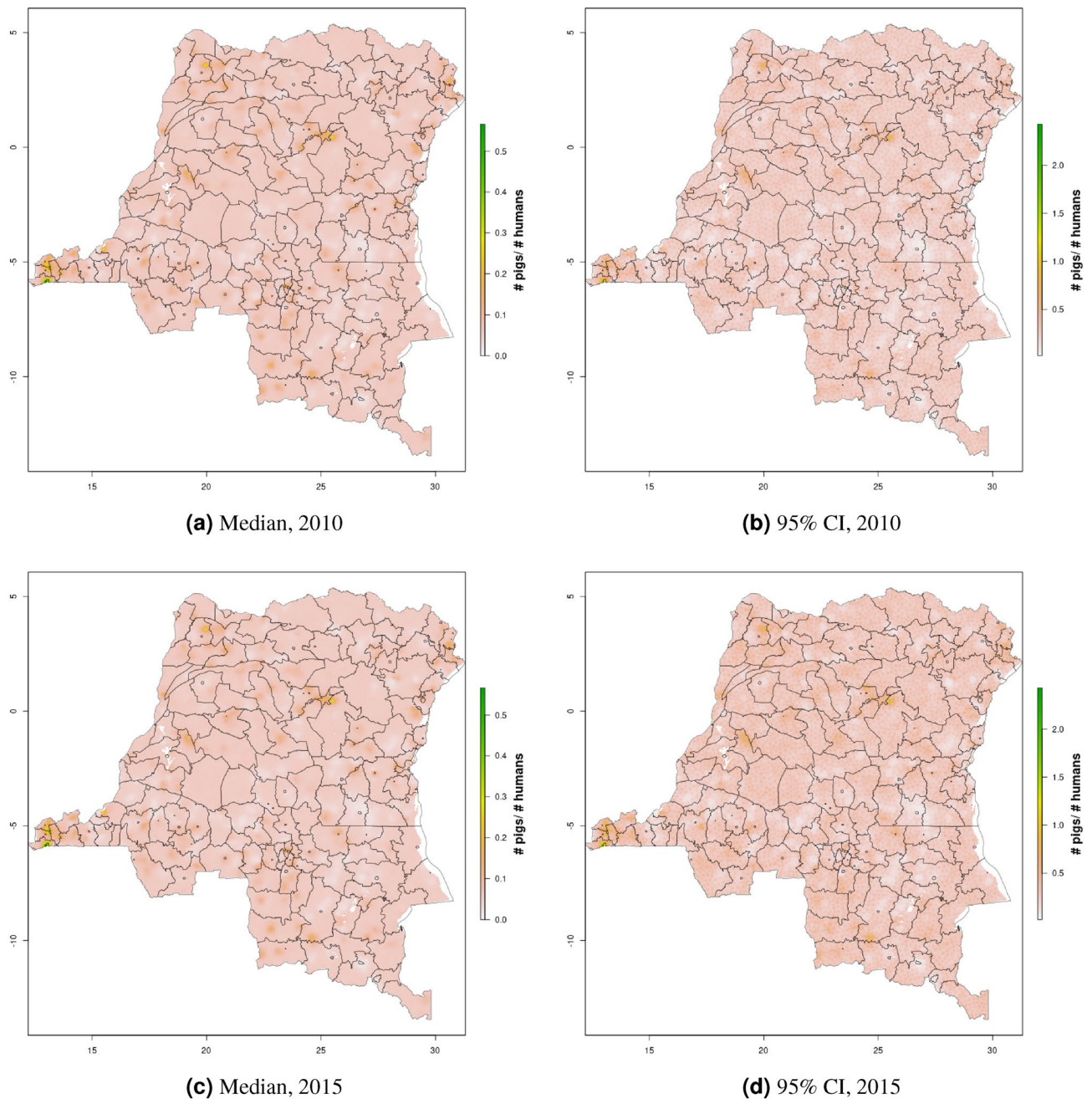
first year was used. These attribute data were then merged with location data. Finally, source-specific data were merged to create one spatial points dataset (Malawi, Uganda, DRC) or spatial polygon dataset (South Sudan), and predictor data were extracted (detailed below).

**Model fitting.** *Malawi, Uganda, and DRC.* For Malawi, Uganda, and DRC, we used the SPDE approach to Gaussian process (GP) modeling. The goal of this approach is to smooth over spatial point data and generate a complete surface.

SPDE approach A spatial process  $S(\mathbf{s})$  is a GP if the joint distribution of  $S(\mathbf{s}_1), \dots, S(\mathbf{s}_n)$  (over the whole study area) is an  $n$  dimensional Gaussian distribution for any integer  $n$  and any set of locations  $\mathbf{s}_i$  (e.g., defined by latitude and longitude). That is, the spatial process at each location has a Gaussian distribution.

In Bayesian hierarchical spatial modeling, spatial dependence is generally represented by a precision matrix (inverse of a variance-covariance matrix), with the Matérn family being a popular choice of covariance function<sup>23</sup>. It can be shown that Matérn GPs can be represented as the solution an SPDE, but for a GP with  $n$  clusters, a  $n$  dimensional normal distribution must be modeled, which quickly becomes computationally intractable. These continuous GPs can be discretized (approximated) via a finite element approach using a weighted combination



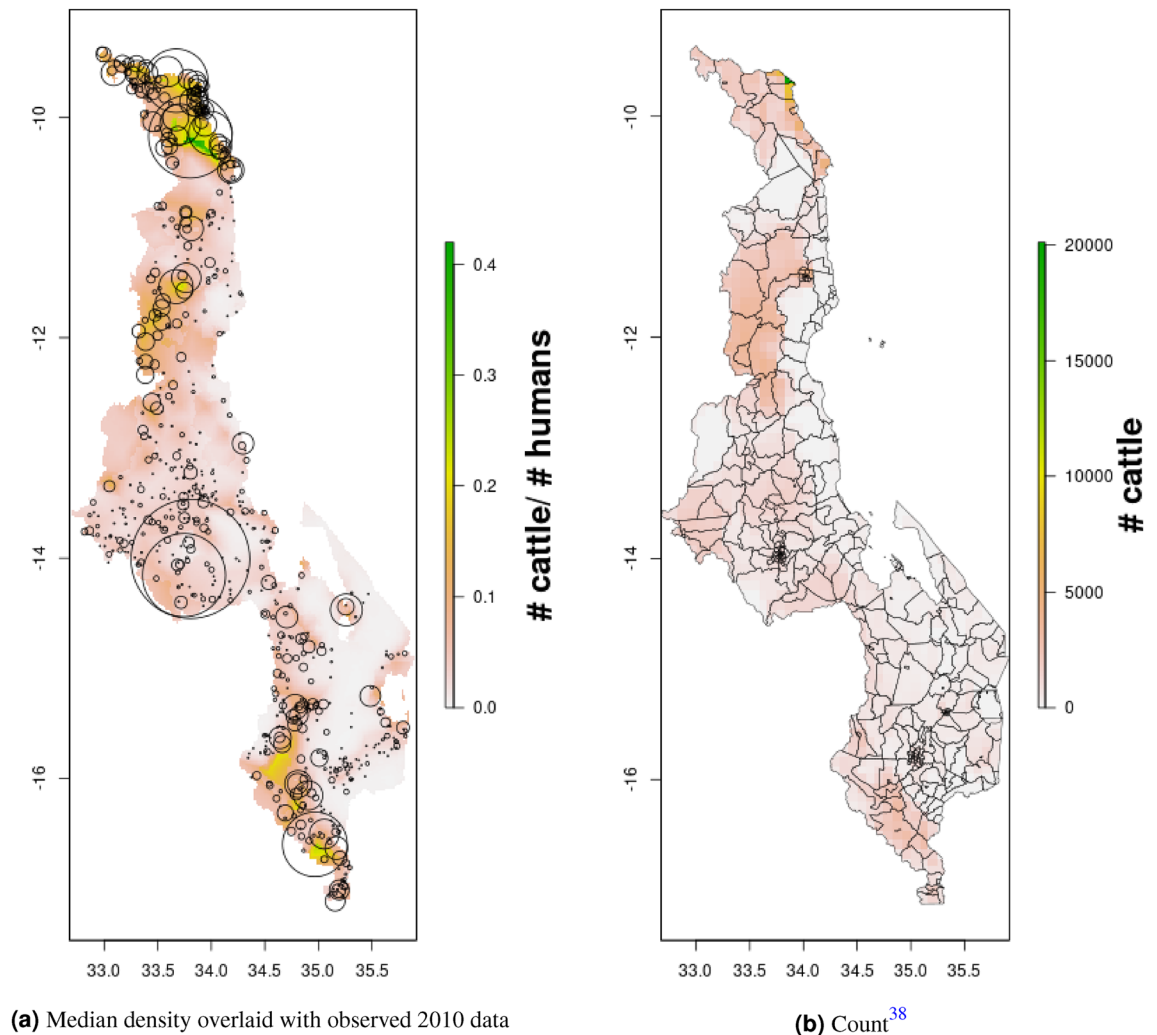


**Figure 9.** Pig density mapping results, DRC, 2010 and 2015. Median (**a, c**) and width of posterior 95% credible interval (**b, d**).

of basis functions. If the weights are multivariate Gaussian and picked to have a sparse precision matrix (many 0 entries), the resulting approximation is a Gaussian Markov Random Field (GMRF). This means a particular location's spatial process depends only on that of its neighbors, not the entire map, markedly simplifying computation<sup>24</sup>. In our application of the SPDE approach, we parameterize the Matérn in terms of three parameters:  $\sigma_s^2$  (spatial variance),  $\rho_s$  (spatial range), and  $\nu$  (shape, or smoothness).

In R-INLA, this discrete approximation is done using a set of non-overlapping triangles which together comprise a mesh over the study region. The basis functions are piecewise linear, equal to 1 on a given vertex and 0 on all other vertices. If one imagines elevating a single mesh vertex, a three-sided pyramid is formed, and the discrete approximation to the GP is a weighted combination of these pyramids. A projection matrix  $A$  projects from the mesh vertices to the  $n$  study locations (here, clusters).

Choice of the mesh dictates the resolution of the spatial effect, with a rule of thumb that features more than two triangles large are resolved well, while features smaller than a triangle will be biased in proportion with the triangle size<sup>23</sup>. Our meshes for Malawi, Uganda, and DRC are presented in Fig. 19 below. In all three countries, we set our mesh to have an inner and outer mesh, moving boundary effects away from the study area (i.e., country



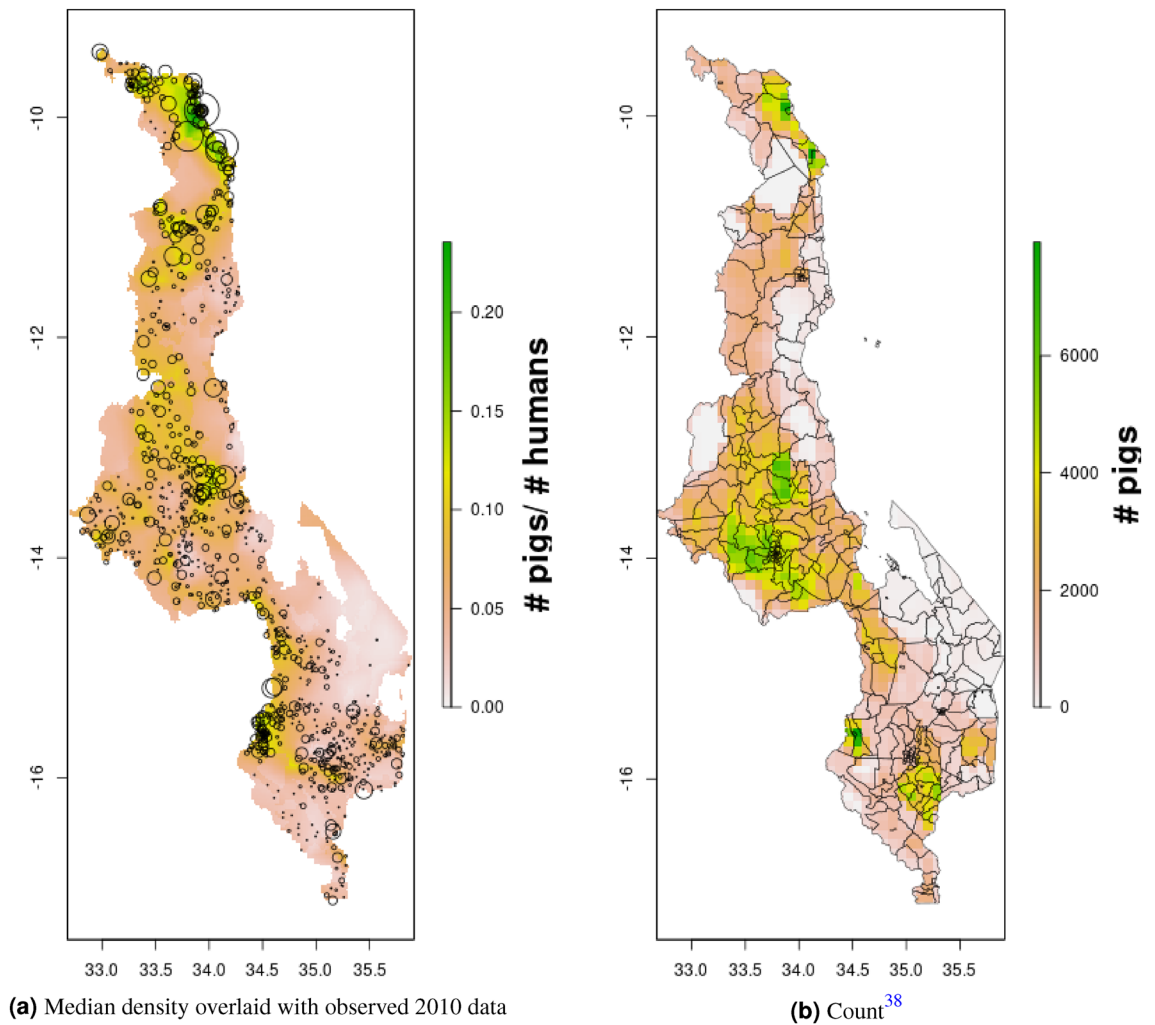
**Figure 10.** Median cattle density (L) compared with estimated cattle counts produced by GLW-3 (R) for Malawi, both 2010. Observed density data for 2010 are presented on the left, indicated by circle size; this figure does not represent the contribution of data from other years to the model for 2010 via the temporal random effect, nor the contribution from predictors<sup>16</sup>.

borders). We set the triangles to be smaller in the inner mesh (maximum edge length of  $0.3^\circ$ ) than the outer mesh (maximum edge length of  $0.6^\circ$ ), and the extension to be  $1^\circ$ .

**Predictors** When the data are complex survey data, the sampling scheme may be accounted for by including design variables—district and urban/rural status in this case—in the regression model. As district is confounded by space, the spatial random effect (detailed above) accounts for this aspect of study design, and district was not included as a predictor in our models.

Predictors included urban/rural status (binary), protected areas (binary), elevation, and bodies of water (binary). While there are many definitions of urbanicity, as urban/rural status is a design variable, the desired definition for this predictor is the value that a given cluster would take had it been sampled by a household survey in a given year. As the data sources we used for livestock mapping largely used the most recent population and housing census as their sampling frame, for each country we generated a  $1\text{km}^2$  urban/rural surface for each sampling frame year. We accomplished this using logistic regression models, with population density from WorldPop<sup>13</sup> and nighttime lights from the National Oceanographic and Atmospheric Administration's Nighttime Lights Time Series<sup>25</sup>, as the predictors. For validation, we split the data into a 2/3 training and 1/3 test set, and evaluated model performance on the test set. We found these models performed very well for rural clusters (95% correct for Uganda, 96% for Malawi and DRC), however performance varied across countries for urban clusters (43% for Uganda, 86% for Malawi, 78% for DRC). As the vast majority of clusters in all three countries are rural, overall performance remained adequate.

Data on protected areas came from the World Database on Protected Areas<sup>26,27</sup>, while this database is updated monthly, there is no publicly-accessible archive. For bodies of water we used levels 1 (lakes  $\geq 50\text{km}^2$  and reservoirs  $\geq 0.5\text{km}^3$ ) and 2 (permanent open water bodies with a surface area of  $\geq 0.1\text{km}^2$ ) data from the Global Lakes and Wetlands Database<sup>28</sup>. Finally, we used 7.5-arc-second data—which have a root mean squared error



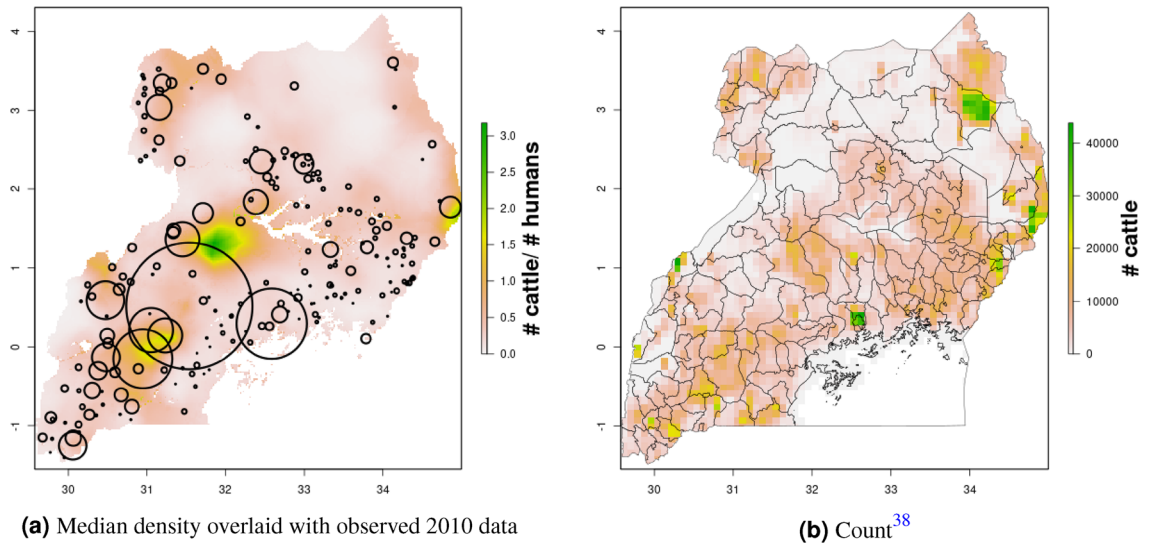
**Figure 11.** Median pig density (L) compared with estimated pig counts produced by GLW-3 (R) for Malawi, both 2010. Observed density data for 2010 are presented on the left, indicated by circle size; this figure does not represent the contribution of data from other years to the model for 2010 via the temporal random effect, nor the contribution from predictors<sup>16</sup>.

of 26–30 meters—and median elevation from GMTED2010 for elevation<sup>29</sup>. Note, in contrast with GLW-3 these predictors were used to improve model fit, not to mask unsuitable pixels<sup>10</sup>.

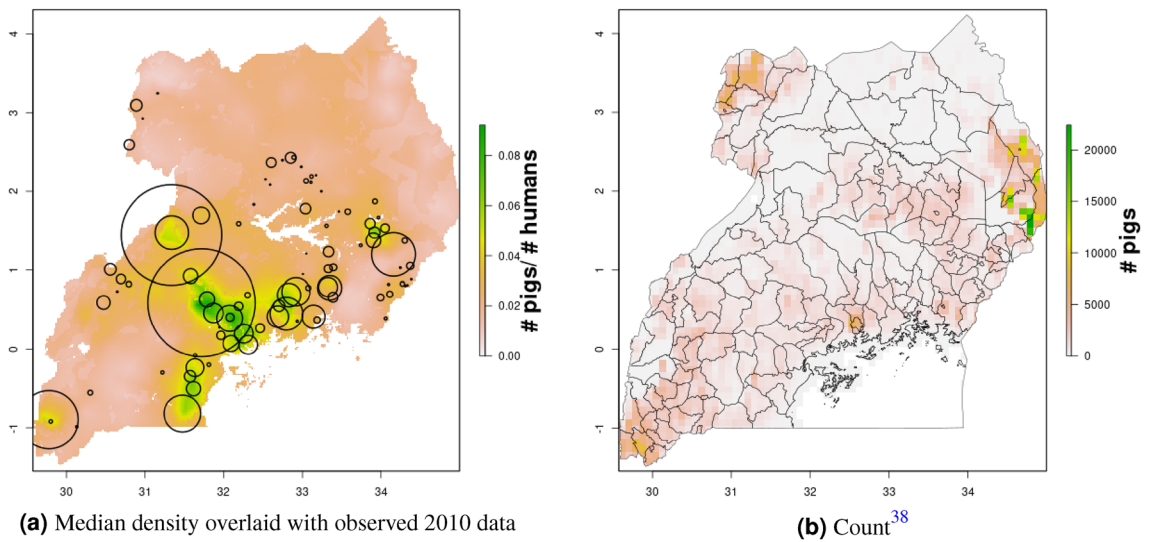
**Models** We fit a three stage Bayesian hierarchical model as follows for the minimal model, fit separately for cattle and pigs, and separately for each country:

$$\begin{aligned}
 Y_i | \mu_i, p_i &\sim \text{ZIP}(\mu_i, p_i) \\
 \mu_i &= \exp\left(\alpha_0 + \beta_1 \mathbf{x}_i + S(\mathbf{s}_i) + \omega_t + \epsilon_i + \gamma_k + \eta_t + u(\mathbf{s}_i, t) + \log(D_i)\right) \\
 S_i | S_j, j \in ne(i) &\sim \text{GMRF}(\sigma_s^2, \rho_s) \\
 \omega_t | \omega_{t-1}, \omega_{t-2}, \omega_{t+1}, \omega_{t+2} &\sim N\left(\frac{4}{6}(\omega_{t-1} + \omega_{t+1}) - \frac{1}{6}(\omega_{t-2} + \omega_{t+2}), \frac{\sigma_\tau^2}{6}\right) \\
 \epsilon_i | \sigma_\epsilon^2 &\sim_{iid} N(0, \sigma_\epsilon^2) \\
 \gamma_k | \sigma_\gamma^2 &\sim_{iid} N(0, \sigma_\gamma^2) \\
 \eta_t | \sigma_\eta^2 &\sim_{iid} N(0, \sigma_\eta^2)
 \end{aligned}$$

where  $i$  indexes cluster,  $p_i$  is the hyperparameter that pertains to the model for the zeroes for cluster  $i$ ,  $Y_i$  is the number of livestock in cluster  $i$ ,  $D_i$  is the number of humans in cluster  $i$  (offset)  $\alpha_0$  is the intercept  $\beta$  is a vector of coefficients for the predictors  $\mathbf{x}_i$  is a vector of predictor variables for location  $i$   $S(\mathbf{s}_i)$  are the spatial random effects (error terms), assumed to follow a Markovian Gaussian random field (GMRF) with variance parameter  $\sigma_s^2$  and range parameter  $\rho_s$ , and  $ne(i)$  are the neighbors of cluster  $i$ .



**Figure 12.** Median pig density (L) compared with estimated pig counts produced by GLW-3 (R) for Uganda, both 2010. Observed density data for 2010 are presented on the left, indicated by circle size; this figure does not represent the contribution of data from other years to the model for 2010 via the temporal random effect, nor the contribution from predictors<sup>16</sup>.



**Figure 13.** Median cattle density (L) compared with estimated cattle counts produced by GLW-3 (R) for DRC, both 2010. Observed density data for 2010 are presented on the left, indicated by circle size; this figure does not represent the contribution of data from other years to the model for 2010 via the temporal random effect, nor the contribution from predictors<sup>16</sup>.

$\omega_t$  is a random walk 2 (RW2) random effect on time (years)  $\epsilon_i$  is an unstructured cluster-level random effect (nugget)  $\gamma_k$  is an unstructured random effect on survey  $\eta_t$  is an unstructured random effect on time (years)  $u(\mathbf{s}_i, t)$  is a space-time interaction which is approximated by an SPDE in space combined with an AR(1) process in time

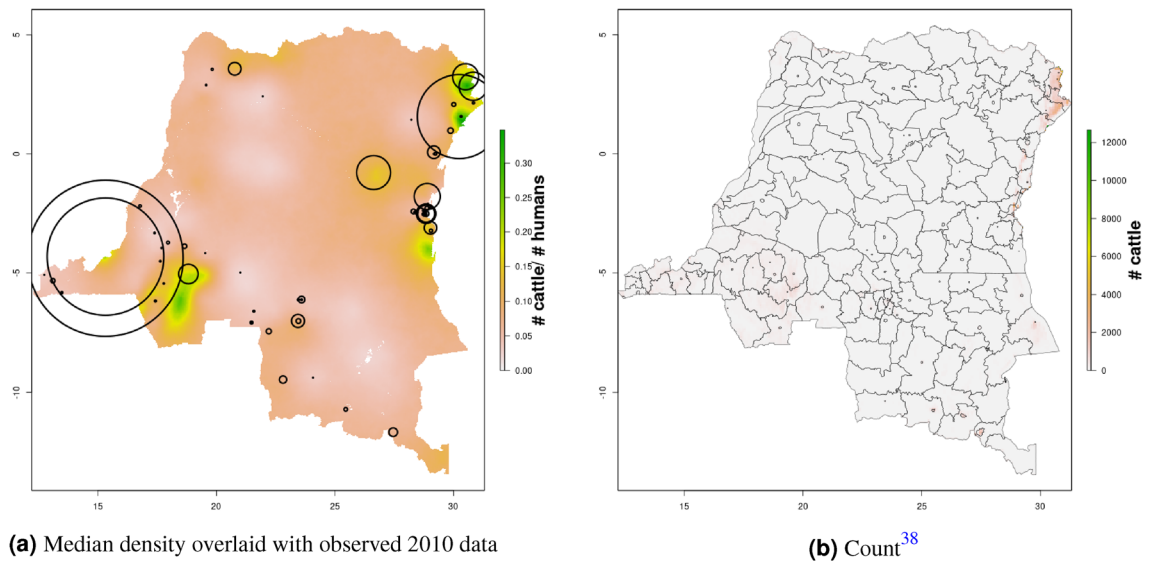
We fit these models as type 1 zero-inflated Poisson (ZIP) models in R-INLA<sup>30</sup>, which combine a distribution for the proportion of zeros with the Poisson distribution. The type 1 likelihood is given as:

$$\text{Prob}(Y|\mu, p) = p \times 1_{y=0} + (1 - p) \times \text{Poisson}(Y|\mu)$$

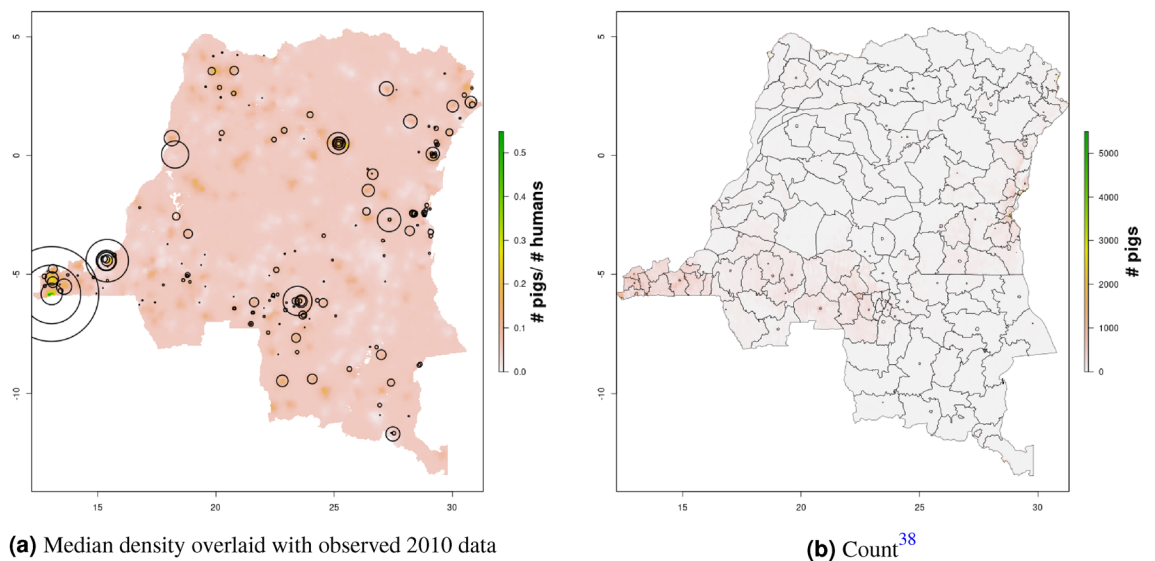
where the first part is the process that generates the zeros (i.e., clusters with no livestock;  $1_{y=0}$  is an indicator that cluster  $i$  has 0 livestock), and the second part generates the livestock counts in cluster  $i$ . This likelihood is a mixture of structural zeros ( $p_i$ ) and sample zeros ( $1 - p_i$ )<sup>31</sup>.

For model selection, we fit a series of five models with equivalent random effects but varying fixed effects: intercept only (model 1); intercept and urban/rural (model 2); intercept urban/rural, and protected area (model 3); intercept, urban/rural, protected area, and bodies of water (model 4); and intercept, urban/rural, protected





**Figure 14.** Median cattle density (L) compared with estimated cattle counts produced by GLW-3 (R) for DRC, both 2010. Observed density data for 2010 are presented on the left, indicated by circle size; this figure does not represent the contribution of data from other years to the model for 2010 via the temporal random effect, nor the contribution from predictors<sup>16</sup>.



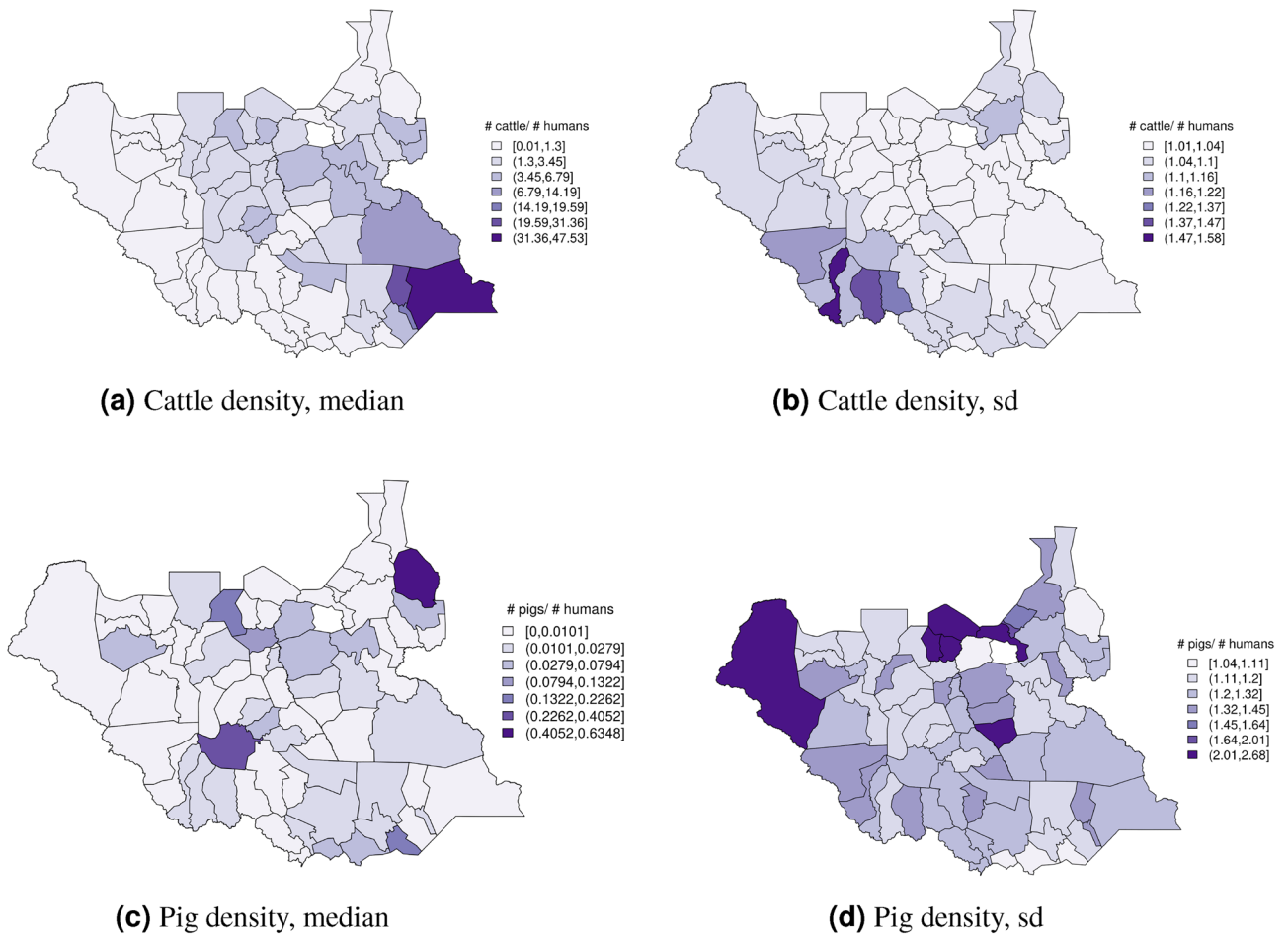
**Figure 15.** Median pig density (L) compared with estimated pig counts produced by GLW-3 (R) for DRC, both 2010. Observed density data for 2010 are presented on the left, indicated by circle size; this figure does not represent the contribution of data from other years to the model for 2010 via the temporal random effect, nor the contribution from predictors<sup>16</sup>.

area, bodies of water, and elevation (model 5). We also fit a sixth model with an RW1 effect in time, rather than RW2, and all fixed effects included in model 5.

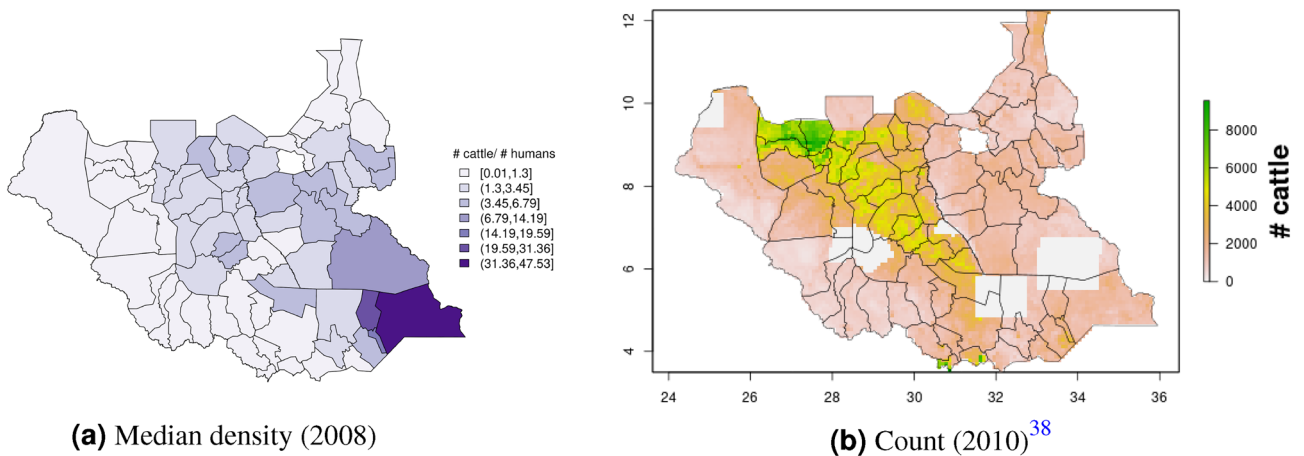
In DRC, due to the limited data availability, we fit models only from 2008-2015 (i.e., extrapolating 2 years in each direction from the input data), and we restricted random effects to the structured (SPDE) effect on space, the structured (RW1 or RW2, depending on the model) effect on time, and the unstructured (iid) effect on space (cluster). That is, the space-time interaction and unstructured random effects for survey and time were removed.

Priors As in Wakefield et al.<sup>32</sup> for the spatial random effect we assigned a fixed shape  $\nu = 1$ , and a “penalized complexity” (PC) prior<sup>33,34</sup> for the spatial range  $\rho_s$  and marginal standard deviation  $\sigma_s$ , such that  $Pr(\rho_s < 0.3) = 0.05$ , and  $Pr(\sigma_s > 1) = 0.05$ . For the spatial range parameter, the selected priors can be interpreted as the 5% quantile corresponds to  $0.3^\circ$ , which is approximately 4% and 9% of the extent of Malawi in the north-south and east-west directions, respectively, and approximately 5% and 1.5% of the extent of Uganda and DRC in each direction, respectively. This spatial range, as well as the mesh, were chosen to be adequately fine to allow for construction of a high-resolution raster map, but adequately coarse to account for jittering of cluster





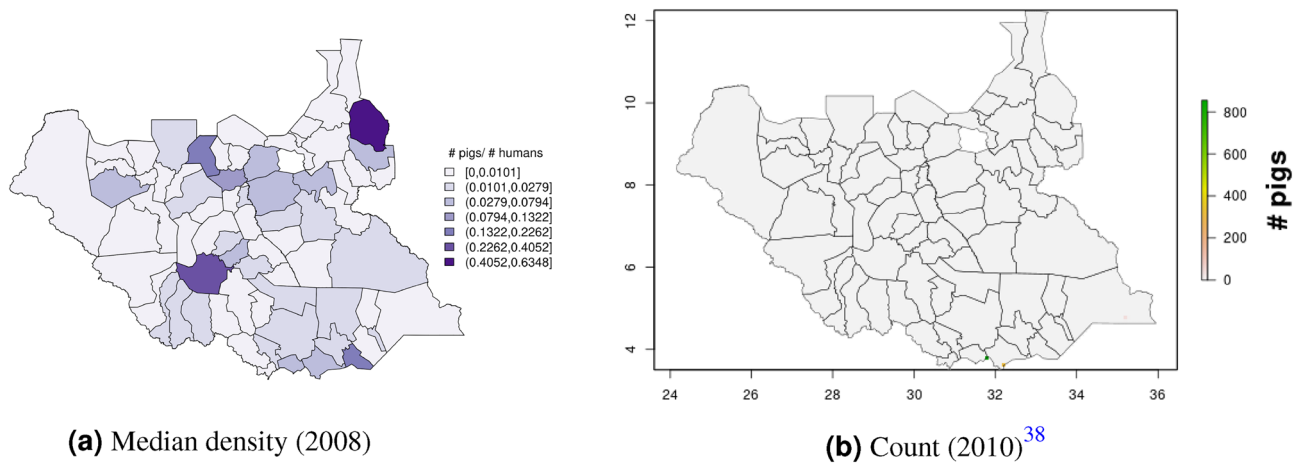
**Figure 16.** Cattle (a-b) and pig (c-d) density mapping results, South Sudan, 2008. Median (a, c) and standard error (b, d).



**Figure 17.** Median cattle density (L, 2008) compared with estimated cattle counts produced by GLW-3 (R, 2010) for South Sudan. Observed density data for 2010 are presented on the left, indicated by circle size; this figure does not represent the contribution of data from other years to the model for 2010 via the temporal random effect, nor the contribution from predictors<sup>16</sup>.

locations in the source data (a means to protect privacy) and to make the model computationally feasible. We used R-INLA's default priors for the structured temporal effects (RW1 and RW2).

For the AR(1) process in time, again using the PC prior specification we set  $Pr(\rho_t > 0.5) = 0.8$  as the prior. Finally, for the precision (inverse of variance) parameters for each of the independent and identically distributed (iid) random effects ( $\sigma_\epsilon^{-2}, \sigma_\gamma^{-2}, \sigma_\eta^{-2}$ ), we specified the prior as  $Pr(\sigma^2 > 0.5) = 0.01$ .



**Figure 18.** Median pig density (L, 2008) compared with estimated pig counts produced by GLW-3 (R, 2010) for South Sudan<sup>16</sup>.

Prediction on a regular grid After fitting our models, we projected livestock density on regular grid, generating  $0.017^\circ \times 0.017^\circ$  grid cells in each country (Fig. 20). This choice governed the resolution of our final raster maps.

We did not use any of the iid (unstructured) random effects for prediction as these were assumed to reflect measurement error. Thus, prediction was comprised of each model's linear predictor, a vector of predictor values for each grid cell, the structured spatiotemporal random effects, and an appropriate projection matrix for each random effect. To produce estimates of uncertainty, we took 1000 draws from the posterior distribution of the structured random effects using the command `inla.posterior.sample()`, and present here the median and width of the posterior 95% credible interval over these 1000 estimates.

Because we did not use an offset in our predictions, we essentially “flattened” the distribution of the human population over the map: our final estimate is that of livestock counts in a grid cell with 1 human being, equivalent to the ratio of livestock to humans, which we are calling livestock “density”.

Model selection and cross-validation After fitting our models, we performed model selection via LOOCV for the cattle maps by survey, using mean squared error (MSE) as our measure of model performance. We defined hold-out sets by survey as we could most readily conceptualize the missing data as hypothetical surveys which were not conducted or samples not collected by a given survey, and we selected MSE due to its interpretability and its ability to capture bias-variance trade-off. Each fold left out a random 25% sample, distributed evenly among sampling strata defined by district and urban-rural status. We calculated MSE as:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_{obs_i} - y_{pred_i})^2$$

where  $y_{obs_i}$  is the observed density at location  $i$ , and  $y_{pred_i}$  is the predicted density at location  $i$ .

External comparisons In addition to cross-validation, we compared our results with two external data sources: GLW-3 (dasymetric product)<sup>35,36</sup> and national-level FAOSTAT total stock estimates<sup>14</sup>. By virtue of the dramatically different scales between our product and these products, these comparisons should be viewed as rough “plausibility checks” rather than formal validation.

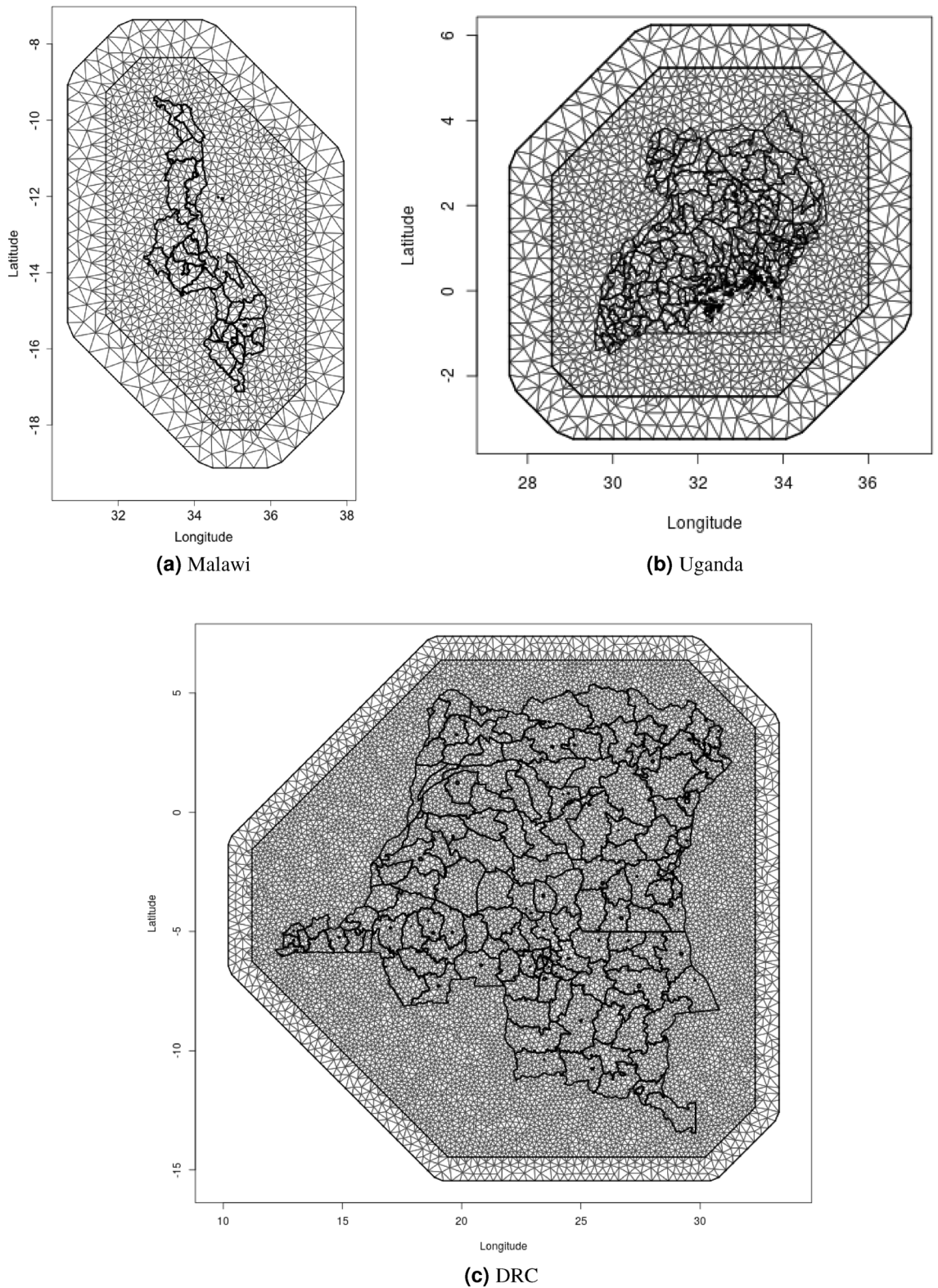
As GLW-3 generated estimates of livestock counts for 2010 only, for this comparison we used WorldPop data as the denominator to derive density estimates<sup>13</sup>, and compared these with our 2010 estimates. We then regressed derived GLW density estimates against our estimates at the pixel-level, using linear regression, and reported coefficient size. For comparison with FAOSTAT total stock data, we used World Bank national (human) population estimates<sup>15</sup> to convert stock counts to density estimates, and compared this with our estimated median density over all years, for each country.

*South Sudan.* Weighted estimates After reading in the data, we used the `svyby()` and `svyratio()` functions in the `survey` package in R<sup>37</sup> to generate weighted estimates of county-level density, using sample weights contained in the IPUMS subset<sup>22</sup> and the ratio estimator below, where  $i$  indexes household and  $c$  indexes county:

$$\hat{Y}_{cHT} = \sum_{i=1}^{n_c} w_{ic} Y_{ic}$$

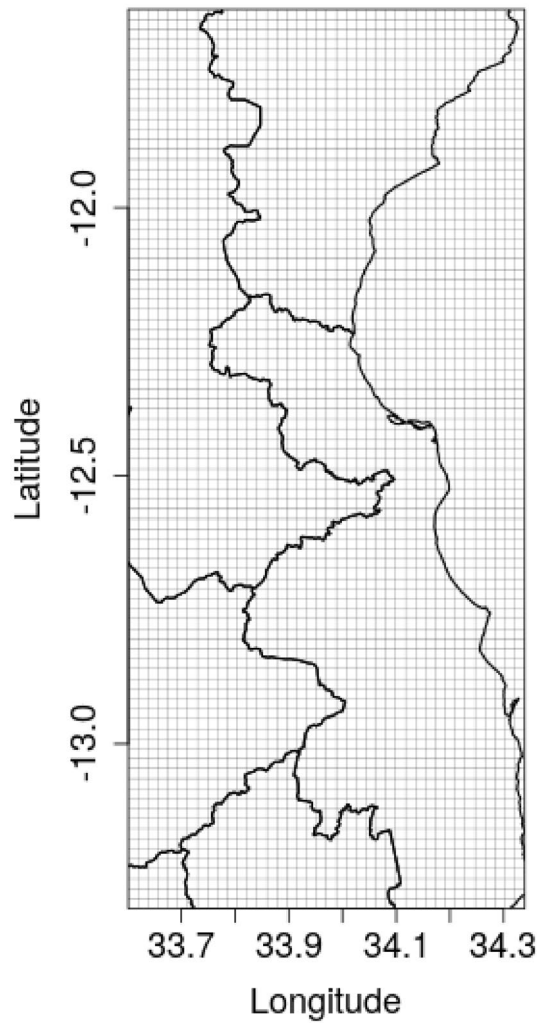
$$\hat{D}_{cHT} = \sum_{i=1}^{n_c} w_{ic} D_{ic}$$

$$\hat{\mu}_c = \frac{\hat{Y}_{cHT}}{\hat{D}_{cHT}}$$



**Figure 19.** Meshes used for SPDE.

where  $Y_{ic}$  is the number of livestock in household  $i$ , county  $c$ , and  $\hat{Y}_{cHT}$  is the Horwitz-Thompson estimate for the number of livestock in county  $c$ .  $D_{ic}$  is the number of residents in household  $i$ , county  $c$ , and  $\hat{D}_{cHT}$  is the Horwitz-Thompson estimate for the (human) population in county  $c$ .  $n_c$  is the total number of households in county  $c$ .  $w_{ic}$  is the sampling weight for household  $i$  in county  $c$ , which is the inverse of sampling probability.



**Figure 20.** Regular grid for prediction, cropped to the northwest of Malawi. Prediction grids were of equivalent resolution in Uganda and DRC.

We also used the `survey` package to estimate design-based standard errors for livestock density. Adapted from Mercer et al.<sup>38</sup> we then specified  $\hat{\theta}_c = \log(\hat{\mu}_c)$ , which, by the delta method, gives us the asymptotic sampling distribution:

$$\hat{\theta}_c | \hat{\mu}_c \sim N\left(\log(\hat{\mu}_c), \frac{\widehat{\text{var}}(\hat{\mu}_c)}{\hat{\mu}_c^2}\right)$$

**Smoothing** After generating the direct estimates ( $\hat{\theta}_c$ ) as detailed above, we next used spatial smoothing to stabilize the variance of  $\hat{\theta}_c$  using a three-stage Bayesian hierarchical model, defined below (note random effects are defined on the transformed scale as in Mercer et al.<sup>38</sup>):

$$\begin{aligned} \hat{\theta}_c | \hat{\mu}_c &\sim N\left(\log(\hat{\mu}_c), \frac{\widehat{\text{var}}(\hat{\mu}_c)}{\hat{\mu}_c^2}\right) \\ \hat{\theta}_c &= \beta_0 + \epsilon_c + S_c \\ \epsilon_c | \sigma_\epsilon^2 &\sim_{iid} N(0, \sigma_\epsilon^2) \\ S_c | S_k, k \in ne(c) &\sim N\left(\bar{S}_k, \frac{\sigma_s^2}{m_c}\right) \end{aligned}$$

where

- $\epsilon_c$  are county-level iid (unstructured) random effects with variance  $\sigma_\epsilon^2$
- $S_c$  are county-level structured random effects which follow the ICAR model with marginal variance  $\sigma_s^2$
- $ne(c)$  denotes neighbors (shared boundary) of county  $c$



Country	Year	Source
Malawi	2004	Second integrated household survey
	2004–2005	Demographic and health survey
	2006	Technology adoption and risk initiative survey
	2010	Demographic and health survey
	2010–2011	Third integrated household survey
	2010, 2013, 2016	Integrated household panel survey
	2012	Malaria indicator survey
	2013–2014	Multiple indicator cluster survey
	2014	Malaria Indicator Survey
	2015–2016	Demographic and health survey
	2016–2017	Fourth integrated household survey
2017	Malaria indicator survey	
South Sudan	2008	Population and housing census
DRC	2010	Multiple indicator cluster survey
	2013–2014	Demographic and health survey
Uganda	2006	Demographic and health survey
	2009	Malaria indicator survey
	2009–2010	National panel survey
	2010–2011	National panel survey
	2011	AIDS indicator survey
	2011	Demographic and health survey
	2011–2012	National panel survey
	2014–2015	Malaria indicator survey
	2016	Demographic and health survey
2018	Malaria indicator survey	

**Table 1.** Data sources by country.

- $m_c$  is the number of neighbors of county  $c$

and ICAR is the intrinsic conditional autoregressive model, which smooths each county's random effect to that of its neighbors, with more smoothing performed for counties with fewer neighbors.

Priors As for the SPDE models, we used PC priors for the smoothing model, with  $Pr(\sigma_\epsilon) > 1 = Pr(\sigma_s) > 1 = 0.01$ . This yields a posterior 95% credible interval for each random effect's residual rate ratio of (0.36, 2.71)<sup>33,34</sup>.

Model selection As there was only one candidate model for South Sudan, model selection was not performed.

**Ethics approval.** As this work relied solely on pre-existing, de-identified data, it does not constitute human or animal subjects research. Per the University of Washington Human Subjects Division, review and approval by the University of Washington Institutional Review Board is not needed (IRB ID: STUDY00004648).

### Data availability

While existing data use agreements do not allow direct sharing of input data, all data were downloaded from publicly-available sources, detailed in Table 1, with links included in our References section. All analyses were performed in R, and all code are available in the GitHub repository linked in the Results section.

Received: 7 January 2022; Accepted: 5 July 2022

Published online: 03 August 2022

### References

1. Herrero, M., Havlik, P., McIntire, J., Palazzo, A. & Valin, H. African livestock futures: Realizing the potential of livestock for food security, poverty reduction and the environment in sub-Saharan Africa. *Office of the Special Representative of the UN Secretary General for Food Security and Nutrition and the United Nations System Influenza Coordination (UNSIC)*, Geneva, Switzerland **188p** (2014).
2. Grace, D. *et al.* Mapping of poverty and likely zoonoses hotspots. *International Livestock Research Institute* (2012).
3. Hermesh, B., Rosenthal, A. & Davidovitch, N. The cycle of distrust in health policy and behavior: Lessons learned from the Negev Bedouin. *PloS One* **15**, e0237734. <https://doi.org/10.1371/journal.pone.0237734> (2020).
4. Hermesh, B., Rosenthal, A. & Davidovitch, N. Boundaries and Politics. *Monash Bioeth. Rev.* **37**, 22–37. <https://doi.org/10.1007/s40592-018-0079-9> (2019).
5. Thornton, P. Livestock production: Recent trends, future prospects. *Phil. Trans. R. Soc. B* **365**, 2853–2867. <https://doi.org/10.1098/rstb.2010.0134> (2010).



6. The Livestock Revolution. <http://www.fao.org/WAIRDOCS/LEAD/X6115E/x6115e03.htm>. Accessed 2 Nov 2018.
7. Steinfeld, H., Gerber, P., Wassenaar, T. D., Castel, V. & de Haan, C. Livestock's long shadow: Environmental issues and options. *FAO* **5**, 7. [https://doi.org/10.1890/1540-9295\(2007\)5\[4:D\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2007)5[4:D]2.0.CO;2) (2006).
8. Lunde, T. & Lintjorn, B. Cattle and climate in Africa: How climate variability has influenced national cattle holdings from 1961–2008. *PeerJ* (2013).
9. Wint, G. & Robinson, T. *Gridded livestock of the world 2007* (Tech. Rep, FAO, Rome, 2007).
10. Gilbert, M. *et al.* Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci. Data* **5**, 1–11. <https://doi.org/10.1038/sdata.2018.227> (2018).
11. Hankerson, B. *et al.* Modeling the spatial distribution of grazing intensity in Kazakhstan. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0210051> (2019).
12. Jahel, C. *et al.* Mapping livestock movements in Sahelian Africa. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-65132-8> (2020).
13. WorldPop: Population/Individual countries 2000–2020. <https://www.worldpop.org/geodata/listing?id=29>. Accessed 25 March 2020.
14. FAOSTAT Live Animals. <http://www.fao.org/faostat/en/#data/QA>. Accessed 16 Feb 2021.
15. The World Bank. Population, total. <https://data.worldbank.org/indicator/SP.POP.TOTL>. Accessed 16 Feb 2021.
16. Gilbert, M. *et al.* Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens, and ducks in 2010. *Sci. Data* <https://doi.org/10.1038/sdata.2018.227> (2018).
17. Demographic and Health Survey: Methodology - collecting geographic data. <https://dhsprogram.com/What-We-Do/GPS-Data-Collection.cfm>. Accessed 22 Oct 2018.
18. European Commission. Global Human Settlement Layer. <https://ghsl.jrc.ec.europa.eu/index.php>. Accessed 09 Apr 2018.
19. Global Health Data Exchange. Institute for Health Metrics and Evaluation. <http://ghdx.healthdata.org/>. Accessed 24 Aug 2017.
20. Central Data Catalog. Integrated Household Survey Network. <https://catalog.ihnsn.org/index.php/catalog>. Accessed 24 Aug 2017.
21. Heger Boyle, M., King, M. & Sobek, M. IPUMS-demographic and health surveys: Version 7 [dataset]. Minnesota Population Center and ICF International, 2019, <https://doi.org/10.18128/D080.V7>.
22. Minnesota Population Center. Integrated public use microdata series, international: Version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019, <https://doi.org/10.18128/D080.V7>.
23. Wakefield, J., Simpson, D. & Godwin, J. Comment: Getting into space with a weight problem. *J. Am. Stat. Assoc.* **111**, 1111–1118. <https://doi.org/10.1080/01621459.2016.1200918> (2016).
24. Diggle, P. & Ribiero, P. Jr. *Model-based geostatistics* (Springer Series in Statistics, New York, 2007).
25. Version 4 DMSP-OLS Nighttime Lights Time Series. National Oceanographic and Atmospheric Administration. <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>. Accessed 30 March 2020.
26. IUCN: World Database on Protected Areas. <https://www.iucn.org/theme/protected-areas/our-work/world-database-protected-areas>. Accessed 09 Apr 2018.
27. Biodiversity A-Z. <https://www.biodiversitya-z.org/content/protected-area>. Accessed 01 Apr 2020.
28. World Wildlife Foundation. Global Lakes and Wetlands Database. <https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database>. Accessed 09 Apr 2018.
29. U.S. Geological Survey. Digital Elevation - Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010). Accessed 01 Apr 2020.
30. Lindgren, F. & Rue, H. Bayesian spatial modelling with R-INLA. *J. Stat. Soft.* **63**, 1–25 (2015).
31. Asmarian, N., Ayatollahi, S., Sharafi, Z. & Zare, N. Bayesian spatial joint model for disease mapping of zero-inflated data with R-INLA: A simulation study and an application to male breast cancer in Iran. *Int. J. Environ. Res. Public Health* **16**, 4460. <https://doi.org/10.3390/ijerph16224460> (2019).
32. Wakefield, J. Multi-level modeling, the ecologic fallacy, and hybrid study designs. *Int. J. Epidemiol.* **38**, 330–336. <https://doi.org/10.1093/ije/dyp179> (2009).
33. Fuglstad, G.-A., Simpson, D., Lindgren, F. & Rue, H. Constructing priors that penalize the complexity of Gaussian random fields. *J. Am. Stat. Assoc.* **114**, 445–452. <https://doi.org/10.1080/01621459.2017.1415907> (2019).
34. Simpson, D., Rue, H., Riebler, A., Martins, T. & Sørbye, S. Penalising model component complexity: A principled, practical approach to constructing priors. *Stat. Sci.* **32**, 1–28. <https://doi.org/10.1214/16-STS576> (2017).
35. Gilbert, M. *et al.* Global cattle distribution in 2010 (5 minutes of arc). <https://doi.org/10.7910/DVN/GIVQ75> (2018).
36. Gilbert, M. *et al.* Global pigs distribution in 2010 (5 minutes of arc). <https://doi.org/10.7910/DVN/33N0JG> (2018).
37. Lumley, T. *survey: Analysis of complex survey samples* (2016). R package version 3.32.
38. Mercer, L., Wakefield, J., Chen, C. & Lumley, T. A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spat. Stat.* **8**, 69–85 (2014).

## Author contributions

J.M.: conceptualization, methodology, software, formal analysis, data curation, writing - original draft, writing - review & editing, visualization. A.K.: data curation, resources, writing - review & editing. M.L.: data curation, resources, writing - review & editing. E.M.M.: data curation, resources, writing - review & editing. A.I.T.: data curation, resources, writing - review & editing. J.W.: methodology, software, writing - review & editing, visualization. A.R.R.: methodology, writing - review & editing. D.P.: methodology, software, writing - review & editing. J.M.: methodology, writing - review & editing. P.R.: methodology, supervision, writing - review & editing, resources.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022