

Improving the accuracy of livestock distribution estimates through spatial interpolation

Ward Bryssinckx¹, Els Ducheyne¹, Bernard Muhwezi², Sunday Godfrey³, Koen Mintiens¹, Herwig Leirs⁴, Guy Hendrickx¹

¹Avia-GIS, Risschotlei 33, BE-2980 Zoersel, Belgium; ²Uganda Bureau of Statistics, P.O. Box 7186, Kampala, Uganda; ³Ministry of Agriculture, Animal Industry and Fisheries, P.O. Box 102, Entebbe, Uganda; ⁴Evolutionary Ecology Group, University of Antwerp, Groenenborgerlaan 171, BE-2020 Antwerp, Belgium

Abstract. Animal distribution maps serve many purposes such as estimating transmission risk of zoonotic pathogens to both animals and humans. The reliability and usability of such maps is highly dependent on the quality of the input data. However, decisions on how to perform livestock surveys are often based on previous work without considering possible consequences. A better understanding of the impact of using different sample designs and processing steps on the accuracy of livestock distribution estimates was acquired through iterative experiments using detailed survey. The importance of sample size, sample design and aggregation is demonstrated and spatial interpolation is presented as a potential way to improve cattle number estimates. As expected, results show that an increasing sample size increased the precision of cattle number estimates but these improvements were mainly seen when the initial sample size was relatively low (e.g. a median relative error decrease of 0.04% per sampled parish for sample sizes below 500 parishes). For higher sample sizes, the added value of further increasing the number of samples declined rapidly (e.g. a median relative error decrease of 0.01% per sampled parish for sample sizes above 500 parishes). When a two-stage stratified sample design was applied to yield more evenly distributed samples, accuracy levels were higher for low sample densities and stabilised at lower sample sizes compared to one-stage stratified sampling. Aggregating the resulting cattle number estimates yielded significantly more accurate results because of averaging under- and over-estimates (e.g. when aggregating cattle number estimates from subcounty to district level, $P < 0.009$ based on a sample of 2,077 parishes using one-stage stratified samples). During aggregation, area-weighted mean values were assigned to higher administrative unit levels. However, when this step is preceded by a spatial interpolation to fill in missing values in non-sampled areas, accuracy is improved remarkably. This counts especially for low sample sizes and spatially even distributed samples (e.g. $P < 0.001$ for a sample of 170 parishes using one-stage stratified sampling and aggregation on district level). Whether the same observations apply on a lower spatial scale should be further investigated.

Keywords: livestock survey, sample size reduction, accuracy, cost efficiency, population estimates.

Introduction

Livestock distribution maps provide a good basis when planning health campaigns and interventions (Mott et al., 1995) as they provide a means for identifying high-risk areas in terms of disease transmission. This is especially the case in regions such as sub-Saharan Africa where environmental factors such as high humidity and high temperatures are ideal for pathogens to survive (Gage et al., 2008) and there is a close interface between livestock and humans (Slingenbergh et al., 2004). In such settings animal welfare also has important economic consequences

altering the livelihood of people living in affected areas (Zinsstag et al., 2007). Accurate methods to estimate the spatio-temporal distribution of livestock are therefore required, especially when detailed data is lacking such as in many countries with extensive systems.

Existing livestock atlases, such as the Global Livestock Production and Health Atlas (Clements et al., 2002) and the Gridded Livestock of the World (Wint and Robinson, 2007) can be used for various decision-making activities but are often based on poor quality livestock input data. High quality field data on the distribution of reservoir species, vectors and hosts is essential when intervention campaigns against infectious diseases are planned (Vaughan and Ormerod, 2003).

In countries with intensive agriculture, livestock owners are required to communicate livestock numbers to the authorities and spatial scale of livestock distribution maps will only be restricted by legal regu-

Corresponding author:
Guy Hendrickx
Avia-GIS, Risschotlei 33, BE-2980 Zoersel, Belgium
Tel./Fax +32 3458 2979
E-mail: ghendrickx@avia-gis.com

lations regarding confidentiality (Bland, 2002). Unfortunately this does not apply to countries with extensive farming systems such as in sub-Saharan Africa where in most countries, livestock statistics are inaccurate, outdated or even absent (Catley, 2006). A series of methods have been developed to estimate livestock densities in these countries ranging from complete censuses to surveys based on a sample design to avoid full household enumeration. The results are then extrapolated to estimate the total number per administrative unit (Wint et al., 2002).

Most sample designs used for national livestock surveys are based on multi-stage sampling because of cost-efficiency. In the study area a number of regions is selected and livestock for each household in the selected regions are enumerated. The primary sampling unit is thus the household as domestic livestock is by definition linked to human presence. Alternatively, only a part of the households within each region may be selected. The selection of sampling regions can be randomly distributed over the entire study area or can be based on a differential selection probability between strata (Stehman and Selkowitz, 2010). The latter provides the opportunity to apply a higher sampling density in regions with a more variable number of livestock or to ensure an even distribution of samples among regions of different categories. The stratification has to be based on a factor expected to influence the livestock enumeration such as vegetation cover.

The methods described above only apply to regions with sedentary husbandry systems, such as mixed farming systems where arable farming and livestock herding are combined. In pastoral systems, where livestock is herded in vast open areas of land during a prolonged period of time, other methods such as aerial survey (Milligan and de Leeuw, 1983) or enumeration in concentration zones near water access points (Michel, 2000) could be applied. Because of their entirely different nature, the evaluation of these methods falls outside the scope of this study.

Geographical information systems (GIS) are often used to select samples, plan survey activities and process the acquired field data (Matthies et al., 2007). However, GIS tools should be data- instead of technology-driven. Since the quality of the input data determines the output quality and given the cost of sampling (Groves, 2004), it is of major importance that survey sample design and processing steps are well adapted to cattle distribution data requirements.

This paper reviews the impact of applying different sample designs and selecting different sample sizes on

output accuracy (the closeness of agreement between the value obtained by using the entire dataset and the result which is obtained by applying the experimental procedure on a sample). Special attention is given to the level of spatial aggregation since the optimal scale on which livestock numbers can be presented is often unknown due to inadequate validation data (Wint and Robinson, 2007). A potential way of increasing accuracy by applying spatial interpolation techniques is presented using Uganda survey data as an example. Issues regarding the efficiency of sampling strategies are also addressed.

Materials and methods

Uganda 2008 national livestock census data

To keep its livestock data up-to-date, the Uganda Bureau of Statistics (UBOS) and the Ministry of Agriculture, Animal Industries and Fisheries (MAAIF) decided to initiate a new national livestock survey in 2008. The previous survey was carried out in 1990/1991 (MAAIF, 1993, 2009). Covering 15.1% of the total number of households, the Uganda 2008 National Livestock Census forms an ideal dataset to test sample designs on.

Uganda is located on the East African plateau, more specifically at the north and north-west of Lake Victoria covering 241,139 km². Neighbouring countries are Sudan, Kenya, Tanzania, Rwanda and the Democratic Republic of the Congo. The highest peak of Mount Stanley is 5,113 m above sea level but the mean altitude within the country is about 1,100 m. A steady slope towards the north is interrupted by Lake Kyoga in the centre of Uganda at an elevation of 914 m. The equatorial climate is moderated by this altitude resulting in varying temperatures between 16 °C and 26 °C between April and November, and over 30 °C during December-March. Rainfall is most abundant in the south (>2,100 mm) and decreases towards the north-east (500 mm). This leads to well-vegetated areas in Western, Central and Eastern regions of Uganda and savannas and dry plains in the Northern region.

As of 2008, Uganda was administratively divided into four regions (Northern, Eastern, Central and Western) and 80 districts (mean surface of 3,000 km²), which are subdivided into counties (mean surface of 1,500 km²), sub-counties (mean surface of 200 km²), and parishes (mean surface of 45 km²) containing several villages. The study area included the entire Ugandan territory. In a first sampling stage, sub-coun-

ties were selected within each district. During the second stage, at least 50 enumeration areas (EA) were sampled from each selected sub-county, in which all households were selected for livestock enumeration (mean number of 109 households per EA). A total of 8,870 EAs and 964,047 households were enumerated between 18 and 25 February 2008. Data on number, breed, age and sex were collected for cattle, small ruminants, pigs and poultry. The questionnaire information was processed at UBOS Entebbe, resulting in a number of datasets. This study focuses on the cattle data.

Based on the individual livestock data entries, the mean number of cattle per household was calculated. To convert mean cattle numbers per household into total cattle numbers per administrative unit, human population data were also needed. The most recent data on human population in Uganda dates from 2002 and was projected to obtain representative numbers for 2008. The methodology was adopted from UBOS and projected human population numbers by fitting a geometric curve to the population census results of 2002 using the following formula (Eq. 1):

$$P_t = P_{0ert} \tag{Eq. 1}$$

where P_t is the projected population at parish level (2008), P_0 is the human population number at parish level from most recent census data, r is the mean growth rate, available on district level and t is the number of years between most recent census and year of projection.

The projected human population size per parish was then proportionally prorated in such a way that the sum of the human population sizes for all parishes within a district would be equal to the projected number stated by UBOS in the State of Uganda Population Report 2008 (Population Secretariat, 2008). This projection was based on the cohort component method as part of the SPECTRUM software and taking (i) age-sex specific survival ratios, (ii) age-specific fertility rates and (iii) net migration into account. Using the processing steps described above obtained the most up-to-date population numbers and allowed comparison of results with national reports.

Livestock population numbers on parish level were obtained using Eq. 2. The mean number of persons per households was available for each district through the 2002 Uganda Population and Housing Census. The resulting livestock numbers were considered as reference to which all sampling methodologies were compared.

$$C_i = C_i^h P_i H_i \tag{Eq. 2}$$

where C_i is the cattle number per parish i , C_i^h is the mean cattle number per household in parish i , i is the admin unit (parish), P_i is the prorated human population size in parish i and H_i is the mean number of persons per household in parish i .

Sample design

To demonstrate the impact of applying different sample designs, two approaches were considered: a one-stage and a two-stage stratified sample design (Fig. 1). The first approach stratified the samples proportionally to district size. The two-stage stratified sample design stratified the samples during the second sampling stage according to five classes of the mean normalized difference vegetation index (NDVI) per parish. The class intervals were determined using natural breaks (Brewer and Pickle, 2002). The sampling of parishes occurred proportional to the total area-size of the parishes belonging to the same NDVI-class in a district. This ensures a true representation of vegetation classes within the sample.

Sample size

Sample sizes of 170, 250, 330, 410, 810, 1,210, 1,610, 2,010 and 2,410 out of 4415 parishes were arbitrarily chosen and covered a range of 4-55% of the total number of parishes in the reference database. The smaller differences between consecutive sample sizes for lower sample sizes were chosen to yield a more detailed description of accuracy changes where large differences were expected. Because the number

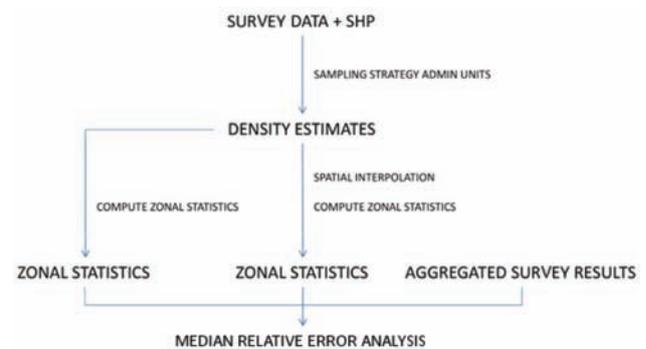


Fig. 1. Experimental setup flow chart based on raw survey data and admin unit geo-references. Density estimates were made on parish level. Zonal statistics for larger admin units were computed using two different approaches (extrapolation and interpolation). The resulting values were compared with aggregated numbers of the entire survey database in a median relative error analysis.

Table 1. Statistical significance testing (extrapolation versus interpolation). Wilcoxon rank sum test results indicate if the added value of interpolation over extrapolation is significant. P-values are given for each of the tested sample sizes, both for one-stage stratified (Fig. 4 and Fig. 5) and two-stage stratified sample designs. For each of the results, the absolute number of sampled parishes is indicated between brackets. The chosen number of samples differs from the final sample sizes due to sample design restrictions. For very large sample sizes, extrapolation and interpolation results are similar when applying one-stage stratified sampling. When two-stage stratified sampling is applied, zonal statistics benefit longer from interpolation techniques.

No. of samples	One-stage stratified		Two-stage stratified	
	P value (no. of samples)	Significance level	P value (no. of samples)	Significance level
170	<0.001 (170)	***	<0.001 (161)	***
250	<0.001 (246)	***	<0.001 (233)	***
330	<0.001 (326)	***	<0.001 (320)	***
410	<0.001 (413)	***	<0.001 (405)	***
810	<0.001 (806)	***	<0.001 (796)	***
1,210	<0.001 (1,163)	***	<0.001 (1,153)	***
1,610	<0.006 (1,487)	**	<0.001 (1,462)	***
2,010	0.438 (1,800)		<0.001 (1,750)	***
2,410	0.617 (2,077)		<0.001 (2,024)	***

Significance codes: 0 “***” 0.001 “**” 0.01 “*” 0.05

of parishes is limited within each district and is not proportional to district size, a large number of samples cannot be allocated entirely. This is why the sample size, given as a parameter for each experiment, may deviate slightly from sample sizes for one-stage and two-stage stratified sampling. Table 1 lists the actual sample sizes. Due to the computational resources needed to run the simulations for each set of parameters, the number of repetitions was limited to 50 repetitions.

The accuracy of cattle density estimates was assessed through a median relative error analysis. First, a relative error value between cattle density estimate and observed cattle density in the reference dataset was calculated for each administrative unit using Eq. 3. This resulted in a median relative error when summarizing the results from every administrative unit within a single iteration. After 50 iterations, an mean median relative error value was calculated. These steps

were repeated for each sample design and administrative unit level (district, county and sub-county) for which cattle number estimates were aggregated.

$$\text{Relative error} = (\text{estimated cattle density} - \text{observed cattle density}) / \text{observed cattle density} \quad (\text{Eq. 3})$$

Spatial interpolation

Generally, an area-weighted mean estimate is extrapolated to larger administrative units when reporting livestock numbers. To test whether an increased accuracy could be obtained for the same sample, a spatial interpolation procedure (Fig. 2) was performed. Polygon-data of estimated cattle densities was first rasterised on a 1 x 1 km pixel resolution. An inverse distance weighted interpolation was used to fill in pixels with no data for non-sampled administrative

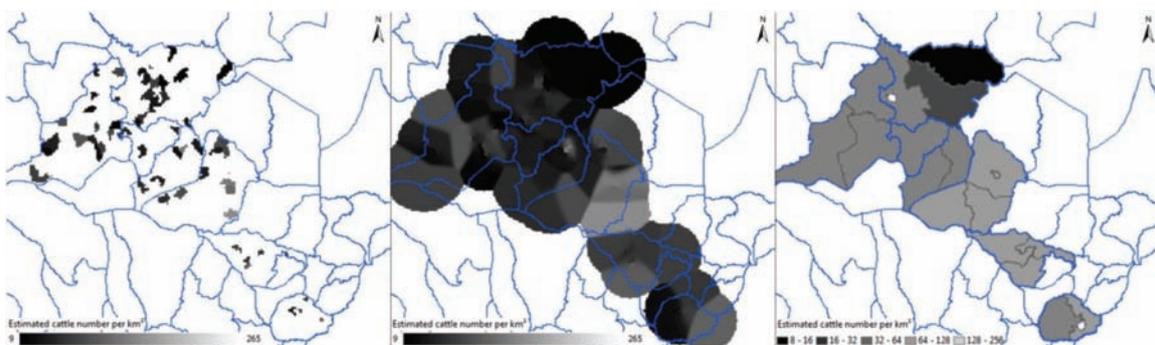


Fig. 2. Spatial interpolation procedure. Fig. 2a represents parishes, which are sampled according to a one-stage stratified sample design (sample design using two-stage stratified sampling is not shown here). For the selected parishes, cattle density numbers are rasterised and interpolated using a Gaussian kernel, which can be seen in Fig. 2b. Fig. 2c shows mean cattle density estimates for each administrative unit (counties in this example) based on pixel values contained by each admin unit. Only admin units, which originally contained sampled parishes were considered.

units. For this purpose, a Gaussian kernel with a radius of 20 km was used as moving window.

Spatial aggregation

Because cattle population sizes were compared at different administrative unit levels (district, county and sub-county) rather than at pixel level, zonal statistics were computed by assigning the mean pixel value as an attribute to the administrative unit. More accurate cattle number estimates were expected after aggregating parish level data to higher administrative unit levels as noise is reduced during this downscaling process.

Due to the interpolation process, additional administrative units were included, depending on the size of the moving window and the kernel shape which were kept constant throughout the experiments. For these administrative units, data only consisted of interpolation results and no original survey results were included. This reduced the reliability of cattle number estimates. Therefore, only regions where parishes were sampled were considered in the analysis.

Comparing accuracy levels

For each experimental run, the median of relative error was recorded to assess general trends. To gain a better understanding of accuracy changes in a broader range of relative errors, third quartile relative errors and maximum relative errors were also computed as these represent the least accurate estimates.

Differences are visualised using box plots. Wilcoxon rank sum tests were used (R-software) to investigate if differences between extrapolated and interpolated cattle numbers were significant at different administrative levels and if spatial aggregation resulted in significant-

ly more accurate estimates. All analyses were performed using R 2.11.1 (R Development Core Team, 2010).

Results

Sample size

The relationship between the number of sampled parishes and the mean median relative error (MRE) is logarithmic (Fig. 3). As an example, for one stage-stratified sampling without interpolation, increasing the sample size resulted in rapid accuracy improvements when starting sample sizes were relatively low: the mean of the MRE decrease when the number of sampled parishes is below 500 is 0.041% per sampled parish. For higher sample sizes the increase in accuracy is much slower: mean MRE decrease of 0.009% per sampled parish.

Spatial interpolation

Sample size did not only affect the accuracy but also the impact of other processing steps on accuracy.

For the one-stage stratified sample design, the difference between MRE of extrapolated and interpolated cattle number estimates is more significant for small sample sizes than for larger sample sizes (Table 1). With a sample size of 170 parishes, there is a significant difference between the extrapolated and interpolated cattle number estimates at the 99% level ($P < 0.001$) (Fig. 4a) while there is no significant difference for a sample size of 2,077 parishes ($P = 0.617$) (Fig. 4c). Third quartile relative errors also show the added value of spatial interpolation over extrapolation (Fig. 5a). Maximum relative errors are lower for interpolation results compared to extrapolation results but

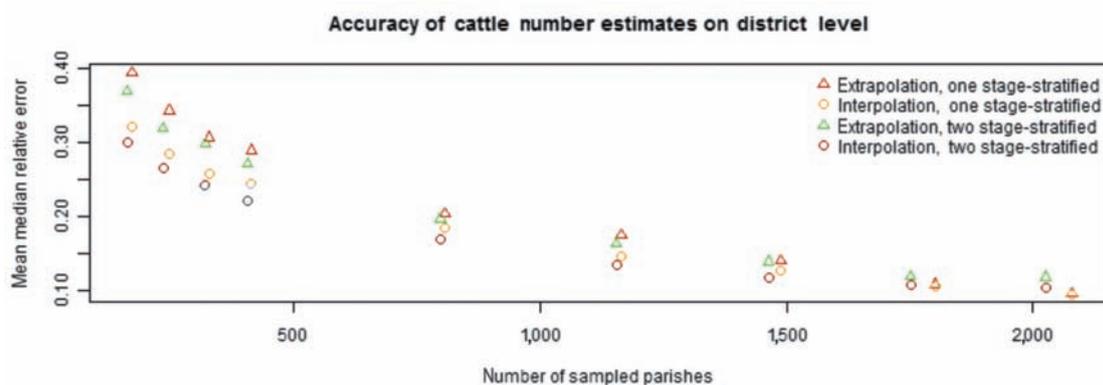


Fig. 3. Accuracy of cattle number estimates. Mean median relative errors representing the accuracy of extrapolation and interpolation results are given for both one-stage and two-stage sample designs. Spatial aggregation was done on district level. The decrease of mean median relative errors when increasing sample size for low sample sizes (e.g. measurements <500) is much more pronounced than for high sample sizes (e.g. measurements >500).

only when considering aggregations to district level (Fig. 5b).

These results are highly dependent on the sample design characteristics. In contrast to the one-stage sampling, the difference between MRE of extrapolat-

ed and interpolated cattle number estimates for the two-stage sampling are significant at larger sample sizes ($P < 0.001$ for sample size of 1,750 samples).

All results show differences between aggregated data at the district level. When methods are compared at the sub-county level, extrapolation outperforms interpolation (Fig. 4c). The same is observed within the set of interpolation methods aggregated at county (Fig. 5b) and sub-county level when only taking large errors into consideration (results not shown). This is due to the higher proportion of parishes sampled at the finer administrative scale.

Sample design

Using the NDVI to categorise parishes into strata in the second sampling stage affected the difference in MRE values between extrapolated and interpolated results (Fig. 3). MRE values for the one-stage stratified approach were higher compared to MRE values for the two-stage stratified approach. This difference became smaller when sample size was increased. When sample size was increased even further, MRE values for two-stage stratified sampling became higher than MRE values for one-stage stratified sampling. Similar results were obtained on different administrative unit levels.

Spatial aggregation

The impact of aggregating cattle number estimates differed when sample design characteristics were changed. For small sample sizes, aggregation from sub-county to a higher administrative unit level had a negative effect on the accuracy (Fig. 4a). The opposite is observed for larger sample sizes (Figs. 4b and 4c). The MRE of spatially aggregated cattle number estimates at a lower scale (district level) is significantly lower than at a finer scale (county level) (e.g. $P < 0.001$ with a sample size of 2,077 samples for extrapolated cattle number estimates using one-stage stratified samples). This is even more pronounced when considering third quartile and maximum relative errors (Figs. 5a and 5b).

Discussion

Sample design

As expected, increasing sample size showed to be an effective way of improving the accuracy of estimates when the initial sample size is relatively low. Other

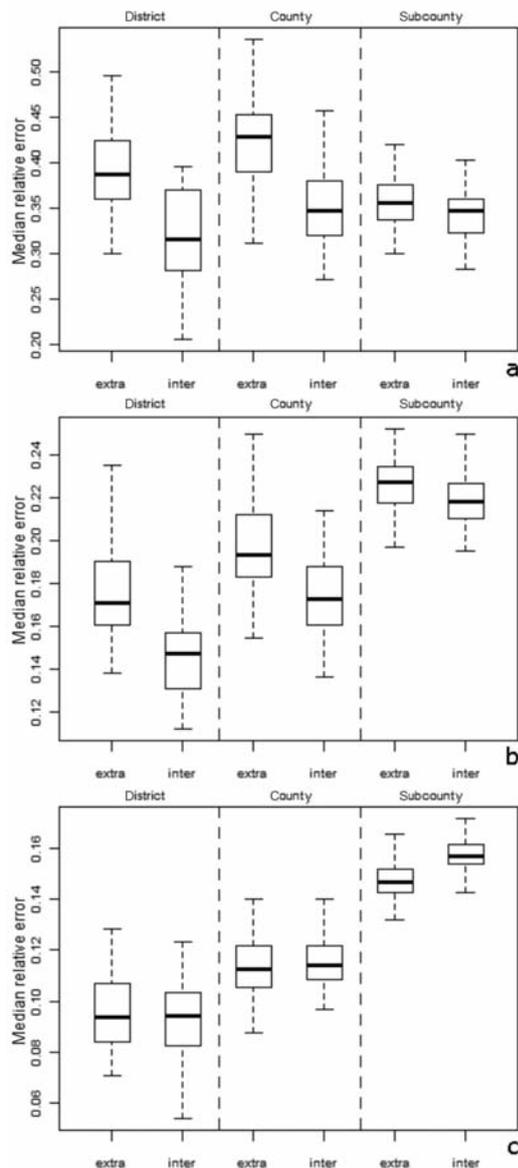


Fig. 4. Median relative error distributions - Median relative error values for extrapolation (extra) and interpolation (inter) procedures show the accuracy of both processing approaches. A comparison is made between district, county and sub-county levels. To represent low, medium and high sample sizes, samples of a: 170, b: 1,163 and c: 2,077 parishes are given as an example. Each time, a one-stage stratified sample design was used. Unless very high sample densities are applied (e.g. Fig. 4c on county and sub-county level), MRE values are lower for interpolation results compared to extrapolation results. Spatial aggregation to higher administrative unit levels also has a clear advantageous impact on accuracy. In Fig. 4a, MRE values are lowest on parish level because of a much higher sample density compared to district and county level. Each time, only admin units, which originally contained sampled parishes were considered.

techniques can also be used to lower the error so that available samples can be used more efficiently. In this study, these techniques were reviewed and tested using the entire dataset as a reference. This was justified by the very large extent of the dataset but it is recognised that derived cattle numbers do not correspond entirely to the real situation.

One way of improving accuracy levels which was verified here, holds the use of prior knowledge to ensure that the sample is representative for the entire area. In this study, this was accomplished by applying a stratified sample design. A first-stage stratification ensured parish selection within districts proportional to district size. A second-stage stratification used NDVI values to categorise parishes and resulted in a

significant accuracy improvement, especially for lower sample sizes. When the number of samples was increased, the probability of obtaining a representative sample from random sample selection increased as well, reducing the added value of applying the two-stage stratified sample design. When sample size was increased further, the second-stage stratification even showed an adverse effect on the accuracy suggesting a more realistic representation from a very large number of random samples compared to a stratification based on one sole environmental parameter.

Spatial interpolation

As cattle presence in extensive mixed-farming livestock systems is strongly related with environmental parameters, changes in cattle density will not appear abruptly but rather gradually just as environmental parameters change. This is why interpolation methods showed to be useful when filling in missing values at non-sampled parishes. However, when most parishes within an administrative unit are sampled, they can also cause adverse effects (Fig. 4c). This especially counts for aggregations on county and sub-county level. In these cases, a simple extrapolation yields very robust estimations, whereas interpolation procedures sometimes result in higher weights for remarkably low or high cattle numbers in neighbouring parishes. Because of few non-sampled parishes, this effect cannot be neutralised by other interpolation results. The added value of interpolation over extrapolation procedures was more pronounced for 3rd quartile relative errors compared to MRE.

For maximum relative errors, interpolation procedures on county and sub-county level resulted in very aberrant values, showing the potential of interpolation procedures to alter cattle number estimations to a large extent. It must be emphasised that these worst-case scenarios are not representative for the general trend.

In these situations, modelling cattle numbers instead of applying simple interpolation techniques might further improve the results. Further investigation is necessary to assess the applicability of statistical models using remote sensing data. Additionally, the effect of interpolation on village level survey results instead of parish level results should be investigated.

Spatial aggregation

The advantage of aggregating livestock numbers to a larger region is already well known and often

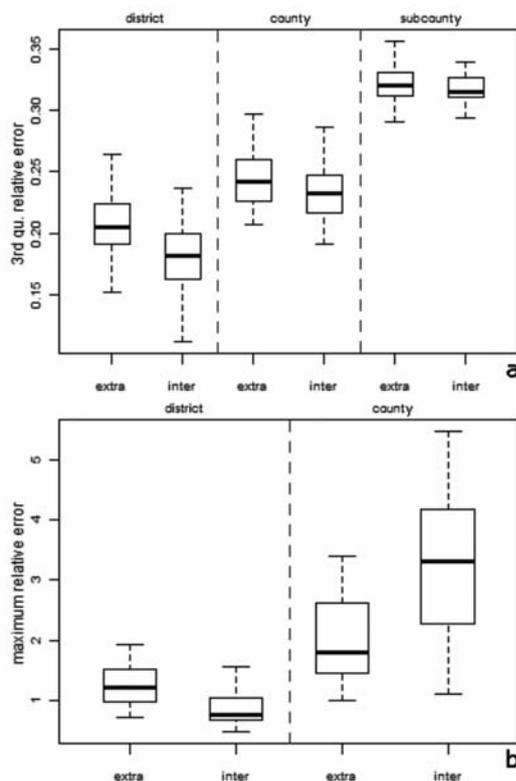


Fig. 5. Third quartile and maximum relative error distributions. 3rd quartile (a) and maximum relative errors (b) of cattle number estimates confirm the added value of interpolation (inter) over extrapolation (extra) procedures. 2,077 parishes were sampled using a one-stage stratified sample design. Results are given on district, county, and sub-county level. Because maximum relative errors for interpolated numbers on sub-county level were very large, they are not shown so that readability stays assured. The reason for high maximum relative error values for interpolations on sub-county level (and to a lesser extent on county level), is the potential of interpolated values to alter zonal statistics strongly when considering relatively small regions. However, maximum relative errors only describe a worst-case scenario and are far from representative for the general tendency.

applied when survey results are processed (Wint and Robinson, 2007). The higher accuracy of aggregated cattle number estimates was also clearly visible in the examples used in this study. For low sample sizes (Fig. 4a), the advantage of aggregating cattle numbers from sub-county level to county level cannot be seen due to the vast difference in sample density between sub-counties and counties. In such a situation, the added value of aggregation is outperformed by the advantage of high sample densities within small administrative units. For high overall sample densities, this difference becomes less pronounced and benefits of spatial aggregation become more visible (Figs. 4b and 4c). Based on MRE ranges in Fig. 4, one could argue that aggregation is superfluous since accuracy differences are small and spatial detail is lost (as can be seen in Fig. 2). However, MRE values only represent a general trend and large differences between administrative unit levels can still be observed when considering relative errors from individual runs or error statistics other than MRE (Figs. 5a and 5b).

The findings regarding the effects of sample design and processing steps on cattle number estimates were based on an iterative series of experiments. Fifty repetitions of each experiment were performed and considered as being capable of demonstrating the response of accuracy levels when sample and processing parameters were changed.

Conclusions

The outcome of sampling techniques depends on many factors and changing one of them will also have an effect on others. Increasing sample size, applying different sample designs (e.g. stratification) and aggregating estimates were already well known to increase accuracy levels. Next to them, spatial interpolation also showed to have a positive effect when processing survey results, unless very large sample sizes are used and spatial aggregation is done to a small administrative unit level. How many samples can be used before interpolation results in less good overall results compared to extrapolation procedures depends on the variability among denominator data in a study area.

Generally, the advantages of interpolation techniques were clear but in extreme cases interpolation can also result in worse estimates, especially when considering maximum relative errors. These extreme cases occur when only few subregions of the administrative unit of interest were sampled (which stresses the importance of a stratified sample design). Whether

this can be improved by using spatial models instead of spatial interpolations should be further investigated. However, even when livestock numbers for as many as 5,669 households are sampled per district (in 40% of all subregions: about 26 out of 67 parishes), spatial interpolation added value by lowering even the highest relative error values. These results show that this approach is a viable method to increase accuracy levels in countries with extensive systems where resources to conduct large-scale livestock surveys are lacking. How vast this increase is, depends on the probability density function of livestock numbers per sample unit. In case of normally distributed denominator data, a small standard deviation will give accurate results by only aggregating the data, leaving less room for improvements. However, as livestock numbers per sample unit are in general not normally distributed, it is difficult to quantify improvements using statistical methods and one should rely on iterative simulations using data from a similar study area, to decide whether or not to apply spatial interpolation techniques.

Acknowledgements

The authors wish to thank the Ministry for Agriculture, Animal Industries and Fisheries (MAAIF) and the Uganda Bureau of Statistics (UBOS) for their kind permission to use the Uganda 2008 livestock census data and Albert Mugenyi for his valuable assistance and contributions. The authors also gratefully acknowledge the funding support from IWT (IWT 90700) and the ICONZ (FP7 221948) project.

References

- Bland WL, 2002. Estimating spatial cattle and swine inventories from limited data. *Agr Sys* 73, 165-172.
- Brewer CA, Pickle L, 2002. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Ann Assoc Am Geogr* 92, 662-681.
- Catley A, 2006. Use of participatory epidemiology to compare the clinical veterinary knowledge of pastoralists and veterinarians in East Africa. *Trop Anim Health Prod* 38, 171-184.
- Clements ACA, Pfeiffer DU, Otte MJ, Morteo K, Chen L, 2002. A Global Livestock Production and Health Atlas (GLiPHA) for interactive presentation, integration and analysis of livestock data. *Prev Vet Med* 56, 19-32.
- Gage KL, Burkot TR, Eisen RJ, Hayes EB, 2008. Climate and vectorborne diseases. *Am J Prev Med* 35, 436-450.
- Groves RM, 2004. Survey errors and survey costs. Wiley-Interscience, 590 pp.
- Henry GT, 1990. Practical sampling. Newbury Park. Sage

- Publications Inc., pp. 144.
- MAAIF, 1993. Report on Uganda national census of agriculture and livestock (1990-1991), Entebbe. Ministry of Agriculture, Animal Industries and Fisheries.
- MAAIF, 2009. The national livestock census report 2008, Kampala. Ministry of Agriculture, Animal Industries and Fisheries.
- Matthies M, Giupponi C, Ostendorf B, 2007. Environmental decision support systems. Current issues, methods and tools. *Environ Modell Softw* 22, 123-127.
- Michel JF, 2000. Mission d'appui au Ministère du Développement Rural du Mali: méthodologie du recensement national du cheptel transhumant et nomade, Rome. Cirad-empt / Agrotec-Spa, n. 01-012. 37 p.
- Milligan K, de Leeuw PN, 1983. Low altitude aerial surveys in pastoral systems research. In: Proceedings, Workshop on Pastoral Systems Research in Sub-Saharan Africa, Addis Ababa. International Livestock Centre for Africa.
- Mott KE, Nuttall I, Desjeux P, Cattland P, 1995. New geographical approaches to control of some parasitic zoonoses. *Bull World Health Organ* 73, 247-257.
- Population Secretariat, 2008. State of Uganda population report 2008, Kampala. UNFPA Uganda, 104 p.
- R Development Core Team, 2010. R: a language and environment for statistical computing. ISBN 3-900051-07-0. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org> (accessed on April 2010).
- Slingenbergh J, Gilbert M, Balogh K, Wint W, 2004. Ecological sources of zoonotic diseases. *Rev Sci Tech Off Int Epiz* 23, 467-484.
- Stehman SV, Selkowitz DJ, 2010. A spatially stratified, multi-stage cluster sampling design for assessing accuracy of the Alaska (USA). National Land Cover Database (NLCD). *Int J Remote Sens* 31, 1877-1896
- Vaughan IP, Ormerod SJ, 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Cons Biol* 17, 1601-1611.
- Wint W, Bourn D, Hendrickx G, Kruska R, Slingenbergh J, 2002. Livestock mapping: present and future. In: Tateishi R, Hastings D (eds), *Global environmental databases, present situation, future directions*. 2nd ed. Chiba, International Society for Photogrammetry and Remote Sensing (ISPRS).
- Wint W, Robinson T, 2007. *Gridded Livestock of the World*. FAO, 131 pp.
- Zinsstag J, Schelling E, Roth F, Bonfoh B, de Savigny D, Tanner M, 2007. Human benefits of animal interventions for zoonoses control. *Emerg Infect Dis* 13, 527-531.