ISSN 2221-8793

9

FAO ANIMAL PRODUCTION AND HEALTH

working paper

POVERTY MAPPING IN UGANDA

Extrapolating household expenditure data using environmental data and regression techniques

Cover photographs:

Left and centre image: ©FAO/Roberto Faidutti
Right image: ©FAO/Giulio Napolitano

9

FAO ANIMAL PRODUCTION AND HEALTH
working paper

POVERTY MAPPING IN UGANDA

Extrapolating household expenditure data
using environmental data and regression techniques

Andrew Nelson
David Rogers
Timothy Robinson

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 2012

Recommended Citation

FAO. 2011. *Poverty mapping in Uganda: Extrapolating household expenditure data using environmental data and regression techniques*, by Nelson, A., Rogers, D.J. & Robinson, T.P. Animal Production and Health Working Paper. No. 9. Rome.

Keywords

Poverty; welfare; livelihoods; mapping; small area mapping; geographic information systems; linear regression; geographically weighted regression; spatial variability; scale; Fourier-processed multi-temporal remotely sensed data.

Authors' details

Andrew Nelson is a geographer with interests in accessibility modelling, population mapping and the spatial analysis of environmental and socio-economic variables.

David Rogers is Professor of Ecology at the University of Oxford, and heads the Trypanosomiasis and Landuse in Africa (TALA) Research Group. He has been responsible for developing the environmental dataset used here, which is based on temporal Fourier-processing of multi-temporal satellite data. This approach reveals the all-important seasonal characteristics of the environment.

Timothy Robinson works for FAO's Livestock Information, Sector Analysis and Policy Branch (AGAL), where he is responsible for the development of livestock information systems. This includes the mapping and analysis of livestock distributions and production systems, and exploring the social, environmental, animal and public health outcomes of the livestock sector.

E-ISBN 978-92-5-107211-0 (PDF)

Contents

Preface

Around 2.6 billion people in the developing world are estimated to have to make a living on less than $2 a day and of these, about 1.4 billion are 'extremely' poor; surviving on less than $1.25 a day. Nearly three quarters of the extremely poor – that is around 1 billion people – live in rural areas and, despite growing urbanization, more than half of the 'dollar-poor' will reside in rural areas until about 2035. Most rural households depend on agriculture as part of their livelihood and livestock commonly form an integral part of their production system. On the other hand, to a large extent driven by increasing per capita incomes, the livestock sector has become one of the fastest developing agricultural sub-sectors, exerting substantial pressure on natural resources as well as on traditional production (and marketing) practices.

In the face of these opposing forces, guiding livestock sector development on a pathway that balances the interests of low and high income households and regions as well as the interest of current and future generations poses a tremendous challenge to policymakers and development practioners. Furthermore, technologies are rapidly changing while at the same time countries are engaging in institutional 'experiments' through planned and un-planned restructuring of their livestock and related industries, making it difficult for anyone to keep abreast with current realities.

This 'Working Paper' series pulls together different strands of work on the wide range of topics covered by the Animal Production and Health Division with the aim of providing 'fresh' information on developments in various regions of the globe, some of which is hoped may contribute to foster sustainable and equitable livestock sector development.

The work described in this paper follows directly on from earlier attempts to develop a novel approach to mapping poverty using environmental data. The aim was to get closer to understanding some of the underlying causes of poverty – something that is unlikely to be feasible using approaches based only on socio-economic data such as the traditional small area estimate (SAE) techniques. The environmental poverty mapping technique involved modelling geo-registered household expenditure estimates in Uganda, available from household surveys, using discriminant analysis of a range of environmental data – mostly derived from satellite remote sensing. This analysis was successful, resulting in a series of poverty maps and lists of environmental variables that were strongly correlated with poverty at different spatial resolutions.

At the time the original analysis the SAE technique had not yet been carried out on the Uganda household survey data for 2002, so no direct comparison of the two approaches was possible. Small area estimate maps for 2002 have since been published by the Uganda Bureau of Statistics (UBOS), so direct comparisons of the environmental techniques against these were now possible.

In the analysis described here we further examined the extent to which environmental data from remote sensing and other sources were correlated with welfare estimates from household survey data. We employed an alternative suite of statistical approaches, compared to the original study, with the ultimate aim of exploring whether different correlates of poverty were important in different parts of the country. As a bench mark, a single Ordinary Least Squares (OLS) regression model was developed for the whole country. The effects of zoning were explored by developing different OLS models for aggregations of households within different livestock production systems. Finally, Geographically Weighted Regression (GWR) was used explicitly to model the spatial variation and scale dependency of the regression coefficients.

The results re-emphasise the important contribution that environmental analysis can make to mapping rural poverty, and ultimately to understanding its distribution and causes.

Acknowledgements

The authors would first like to acknowledge the support and collaboration offered by John Male-Mukasa, Executive Director of the Uganda Bureau of Statistics (UBOS), for his on-going support, and Thomas Emwanu, who provided all of the household survey data used in this analysis. We would also like to acknowledge Francesca Pozzi, Federica Chiozza, William Wint and Claudia Pittiglio who all have made significant contributions to the preparation of much of the spatial data used in the analysis. Claudia Ciarlantini and Carmen Hopmans are acknowledged for the design, layout and formatting of the working paper.

Abbreviations

AGAL	FAO Livestock Information, Sector Analysis and Policy Branch
AIC	Akaike's Information Criterion
AICc	Corrected Akaike's Information Criterion
ANOVA	Analysis Of Variance
AVHRR	Advanced Very High Resolution Radiometer
CCD	Cold Cloud Duration
CSI-SRTM	Consortium for Spatial Information - Shuttle Radar Topography Mission
CV	Cross-Validation
DEM	Digital Elevation Model
ERGO	Environmental Research Group Oxford
FAO	Food and Agriculture Organisation
GAM	Generalised Additive Model
GIS	Geographic Information System
GLW	Gridded Livestock of the World
GPW	Gridded Population of the World
GRUMP	Global Rural and Urban Mapping Project
GWR	Geographically Weighted Regression
IGAD	Inter-Governmental Authority on Development
IGAD LPI	IGAD Livestock Policy Initiative
ILRI	International Livestock Research Institute
LGP	Length of Growing Period
LST	Land Surface Temperature
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MIR	Middle Infra-Red
NASA	National Aeronautics and Space Administration
NDVI	Normalised Difference Vegetation Index
OLS	Ordinary Least Squares
PPLPI	Pro-Poor Livestock Policy Initiative
RMSE	Root Mean Square Error
SAE	Small Area Estimate
Tair	Air Temperature
TALA	Trypanosomaiasis And Land-use in Africa
UBOS	Uganda Bureau of Statistics
UNHS	Uganda National Household Survey
VIF	Variance Inflation Factors

Executive summary

In 2006 the Food and Agriculture Organization (FAO) Pro-Poor Livestock Policy Initiative (PPLPI) published results from the development of a novel approach to poverty mapping in which household survey data were analysed using a suite of environmental variables (Rogers *et al.* 2006; Robinson *et al.* 2007). Discriminant analysis was used successfully to model household expenditure, resulting in a series of poverty maps of Uganda and lists of environmental variables that were strongly correlated with poverty at different spatial resolutions. The spatial data used to model poverty included direct measures of key climatic variables (such as temperature), descriptor variables of key ingredients of poverty-related processes (such as agricultural potential, agricultural production systems and access to markets and services) and proxies for constraints on the health of people, crops and livestock. Whilst such an analysis cannot provide conclusive evidence as to the causes of poverty, it certainly highlights environmental factors that are strongly associated with it.

In this analysis the extent to which spatial data from remote sensing and other sources (which act as proxies for environmental conditions) are correlated with household survey data on expenditure is further examined. For each rural household in the 2002 Uganda National Household Survey (UNHS-2) values from a subset of environmental variables were extracted, based on the results of earlier studies. Averaging data up to a series of different spatial resolutions the spatial variation and scale dependency of regression coefficients was model using: (i) Ordinary Least Squares (OLS) regression for the whole country, (ii) a regional approach based on a different OLS model for different aggregations of a map of livestock production systems, and (iii) Geographically Weighted Regression (GWR).

The model results were compared with each other at a range of pixel resolutions, and also with the Small Area Estimate (SAE) maps derived from the same household survey data.

Across Uganda, vapour pressure deficit (VPD) (negative) and population density (positive) were the two most influential factors associated with (i.e. predictive of) the level of rural household (expenditure). When this was broken down into livestock production systems, the Normalised Difference Vegetation Index (positive) and access to markets (negative) were the most influential in the arid and semi-arid systems and cattle (positive) and VPD (negative) were the most influential in livestock-only systems.

Comparison of these environmental regression models of poverty with the SAE poverty maps revealed that all such models had lower errors than the SAE model. The GWR performed better than the regional OLS, which, in turn, performed better than the country-wide OLS. GWR in particular was able to generate higher resolution estimates of poverty with comparable errors to the much coarser SAE model, offering a seven-fold increase in spatial resolution.

There was significant spatial variation in the GWR regression which did not match the zonation offered by the livestock production systems, suggesting that alternative zonings should be explored in future when developing regional regression models.

Environmental approaches clearly have a role to play alongside more traditional econometric poverty mapping methods, and there is scope to combine the two to explain better the linkages between poverty and the environment and to develop spatial models for more accurate poverty mapping.

Introduction

Rogers *et al.* (2006) showed how the spatial pattern of household expenditure in Uganda could be described in terms of environmental data derived largely from remote sensing. This approach differs considerably from the Small Area Estimate (SAE) technique (Hentschel *et al.* 2000; Elbers and Lanjouw 2000; World Bank 2000), which exploits correlations between socio-economic data collected from detailed household surveys and those collected from housing and population censuses. In the environmental approach the same household survey data are linked, not to census data but instead to environmental data, which are more likely to be representative of factors that cause poverty (agricultural potential, for example) than those that merely reflect it (type of housing, for example). If the sole objective were to map poverty in greater detail than that offered by the household surveys, then the only important criterion by which to compare approaches would be the spatial resolution and accuracy of the predictions. If, however, our objectives go beyond describing the distribution of poverty, and look towards explaining that distribution, then we must turn to more fundamental explanatory variables in our statistical analyses. The underlying assumption in the environmental approach is that people are poor because their environments fail, in some way, to provide the goods and services that are available to those who are less poor.

Such an environmental approach can only be relevant where people are likely to be largely dependent for their livelihoods and welfare on the environment close to where they live, such as in subsistence agricultural systems, where external inputs are minimal or lacking. In Uganda, where 88 percent of the population live in rural areas, the agricultural sector is the main source of livelihoods, employment (73 percent of all employment in 2005/06), and food security (Fan *et al.* 2004); most industries and services in the country are dependent on it (UBOS 2009). Throughout the country, smallholder production predominates, with bananas, cereals, root crops and oil seeds being the main food crops. Some 40 percent of the rural population lives below the poverty line, accounting for 95 percent of the poor in the country as a whole. It follows, therefore, that Uganda provides an appropriate context for environmental poverty mapping, an idea borne out by the results presented in Rogers *et al.* (2006) and Robinson *et al.* (2007).

In the present analysis the extent to which spatial data from remote sensing and other sources are correlated with household survey data is further examined, with the ultimate aim of exploring whether the correlates of poverty vary in different parts of the country. For each geo-referenced rural household with expenditure data in the 2002 household survey corresponding values were extracted from a suite of environmental variables, whose selection was based on the results of previous studies (Rogers *et al.* 2006; Robinson *et al.* 2007; Pozzi and Robinson 2008; Pozzi *et al.* 2009; Rogers *et al.* 2011), and on exploratory analyses described here.

As a bench mark a single Ordinary Least Squares (OLS) regression model was developed for the entire country. The effects of zoning were explored by developing different OLS models for different aggregations of livestock production systems. Finally, Geographically Weighted Regression (GWR) was used explicitly to model the spatial variation and scale dependency of the regression coefficients.

In the following section, below, some of the factors associated with rural poverty in Uganda are discussed. There follows a description of the datasets and statistical methods used. The results are used to explore issues of scale and methodological accuracy and are compared with the SAE results using the same household data. The final section draws conclusions and suggests how the environmental approach to poverty mapping may be taken forward.

Determinants of rural poverty in Uganda

Past research has identified geographical, historical, biophysical and economic factors that influence rural poverty in Uganda. The most frequently quoted factors are natural resources, farming systems, access to markets and infrastructure, and population density. These factors and their relevance are briefly reviewed below, and potential variables or proxies for these factors are presented.

NATURAL RESOURCES

Human survival depends on natural resources which are turned, by agricultural and industrial activities, into goods and services for the maintenance of human communities, their welfare and economic development. In agriculture-dependent subsistence communities, poverty levels are likely to depend on a number of factors that can affect agricultural productivity, including;

- Climate variables, such as temperature and rainfall.
- Length of the growing period (LGP).
- Vegetation activity and phenology indicators, such as multi-temporal vegetation indices.
- Terrain characteristics, such as slope and elevation.
- Soil quality indicators, such as measures of physical and chemical soil properties.

These factors can act and interact in complex ways. For example, acid soils, are favourable for coffee production but not for maize and bean production. Abundant rainfall can promote crop growth and high yields, but may also favour crop pests. Poor human nutrition leads to lower levels of human health, affecting future productivity.

FARMING SYSTEMS

Agricultural activities are the largest source of income in rural Uganda. Greater levels of crop and livestock production and greater ownership of land and livestock assets usually suggest greater levels of affluence. A predominance of livestock, however, can also occur in poor communities where livestock is the only livelihood option. For example, poor pastoralists are isolated and live a nomadic existence that is heavily dependent on ruminant livestock. Ownership of mainly monogastric species in a land-less peri-urban context is not indicative of wealth but a reflection of poverty. Since this study focuses on livestock production systems, possible relevant variables from agricultural census data include the densities of cattle, sheep, goats, pigs and poultry. Ideally, however, data on livestock ownership and the contribution made by livestock to peoples' livelihoods should also be included.

ACCESS TO MARKETS AND INFRASTRUCTURE

Von Thünen was a nineteenth century economist and landowner in North Germany who developed a theory of land-use patterns based on the marginal productivity of land at different distances from a major city in which (it was assumed) all the productivity was sold. In this theory, different types of agricultural production systems would be most profitable at different distances from the city (e.g. dairying and intensive farming nearest to the city and ranching farthest from it). Von Thünen directly and indirectly provided theories on pricing, land use intensity, specialisation and economies of scale (Garnick 1990; Chomitz and Gray 1996). Key variables to take account of these ideas therefore include distances and time taken to travel to roads, centres of population, markets and agricultural inputs such as labour, animal health services and feed (in the case of livestock farming).

POPULATION DENSITY

Areas of high productivity tend to have high population densities. Higher densities of people also imply greater labour availability and greater consumer demand. Rural population densities increase in the vicinity of urban areas and close to transport networks, and are naturally correlated with access to markets.

HEALTH OF PEOPLE, CROPS AND LIVESTOCK

The prevalence of diseases - in crops, livestock and people - is also key to welfare. Some of this is direct; human health is itself a measure of welfare and dealing with human health problems and controlling diseases in crops and livestock are often major expenses in poor households, for example. Other effects are indirect; human ill-health impacts on agricultural labour productivity, for example. Whilst explicit data on these issues may not be available at an appropriate scale and resolution for Uganda, there are often strong correlations between disease prevalence or vector abundance and similar environmental variables to those used here (see for example the reviews in Hay *et al.* 2000 and Pfeiffer *et al.* 2008). These remotely sensed environmental variables thus have the potential to capture much of the variability in the health of people, and their crops and livestock.

OTHER FACTORS

Many other factors have been shown to relate to rural poverty and agricultural productivity. Land tenure is one of the most important whereby greater land security is thought to lead to greater output and better land management practices. Access to credit (banks, rural credit systems and micro credit) can make a difference if directed at small holders, as can the provision of extension services, adoption of new products and technology and the capacity to innovate. Unfortunately, such variables were not available for Uganda in sufficient detail for their inclusion in the present analysis. Here, the emphasis is on remotely sensed and other environmental datasets to provide independent variables for the analyses, resulting in a strong environmental bias, as opposed to the more prevalent socio-economic approaches.

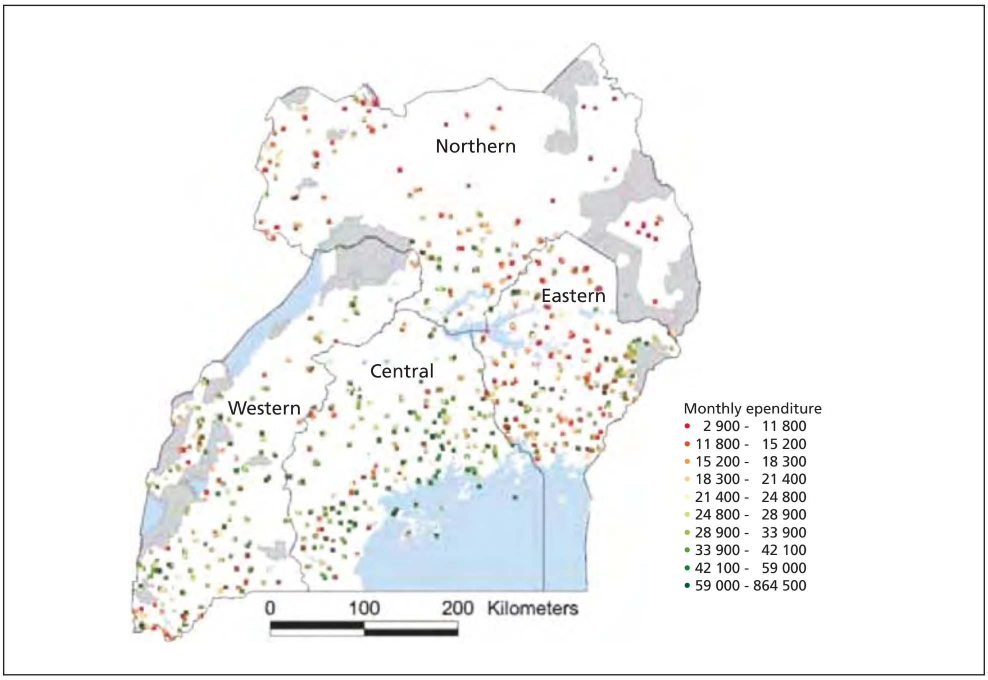
Data

This section describes briefly the data used in the present analysis. Unless otherwise stated, all spatial data are stored as ESRI shape files (points, lines and polygons) or ESRI grids (raster) in geographical co-ordinates (Uganda straddles the equator, so scale distortions are minimal). Map legends for expenditure are based on deciles computed from the household level or aggregated (at about 1km) household level estimates. Grey shading indicates protected areas on the maps.

HOUSEHOLD SURVEY DATA

The Uganda Bureau of Statistics (UBOS) has carried out a number of nationally representative surveys since 1988 (see Table 1 in Rogers *et al.* 2006). In this analysis the second Uganda National Household Survey (UNHS-2) was used, which was carried out between May 2002 and April 2003 (UBOS 2003). Data for 5 614 rural households with reliable geographical coordinate data were selected from a total of 9 711 records (urban and rural) in the survey. The dependent variable used was monthly household expenditure, corrected for the number of adult equivalents per household. Figure 1 shows the location of the households and Table 1 provides summary statistics for the regional differences in rural per-adult equivalent monthly expenditure in Ugandan Shillings. The monthly expenditure data did not exhibit a normal distribution so were transformed before prior to the analysis, as described below.

Figure 1. Rural household locations from the 2002-3003 Uganda National Household survey, showing monthly adult equivalent expenditure (in Uganda shillings).



Note: The administrative boundaries shown refer to the four regions of the country.

There are clear regional differences, with the Central and Western regions having higher levels of expenditure and correspondingly lower percentages of households below the poverty line than the Eastern and Northern regions. Across Uganda, 38 percent of the rural households in the survey were below the poverty line, but this varies from 24 percent in the Central region to 60 percent in the Northern region.

Table 1. Descriptive statistics for rural, monthly adult equivalent expenditure 2002-2003.

a) Summary statistics

Region	Count	Mean	Std Err Mean	Std Dev	Skewness	Kurtosis
Uganda	5 614	32 492	417	31 255	7.6	130.2
Central	1 515	41 153	1 009	39 286	8.1	135.4
Western	1 479	34 237	711	27 332	3.6	22.8
Eastern	1 563	28 813	754	29 816	9.0	131.5
Northern	1 057	23 074	614	19 962	4.8	45.0

b) Quartiles including the Inter Quartile Range (Upper – Lower Quartile)

Region	Minimum	Lwr Q	Median	Upr Q	Maximum	IQ Range
Uganda	2 915	16 728	24 813	37 377	864 534	20 649
Central	4 752	21 858	31 140	47 200	864 534	25 342
Western	3 556	18 382	26 910	40 002	349 200	21 620
Eastern	4 444	15 595	22 681	32 809	608 589	17 215
Northern	2 915	12 102	18 138	26 722	304 400	14 620

c) Poverty lines and rates

Region	Poverty line	Above	Below	Above %	Below %
Uganda	20 760	3 466	2 148	62%	38%
Central	21 322	1 156	359	76%	24%
Western	20 308	1 010	469	68%	32%
Eastern	20 652	875	688	56%	44%
Northern	20 872	425	632	40%	60%

SMALL AREA ESTIMATE POVERTY DATA

Whilst various methods have been used for poverty mapping, some reviewed by Davis (2003), the most common is the SAE technique, discussed by Ghosh and Rao (1994) and developed and exemplified in a series of World Bank studies (*e.g.* Hentschel *et al.* 2000; Elbers and Lanjouw 2000; World Bank 2000). This involves the application of econometric techniques to combine sample survey data with census data to predict poverty indicators using all households covered by the census. The survey provides the specific poverty indicator and the parameters, based on regression models, to predict the poverty levels for the census households. Usually the poverty indicator is a consumption- or expenditure-based indicator of welfare, such as the proportion of households that fall below a certain expenditure level (*i.e.* the poverty line). The basic methodology is quite simple. At the 'zero stage' the comparability of data sources is established and variables common to the census and survey are identified. In the 'first stage' a regression model is estimated between log per capita consumption or expenditure in the household survey and the variables common both to survey and census. The model thus provides a set of empirical regression parameters. These regressions are generally nested at various spatial levels, from regional down to household levels. In the 'second stage' these regression parameters are applied to the census households, where they are used to predict consumption or expenditure in the much more extensive census population, and thus to estimate poverty and inequality for each group of interest. The precision of the poverty estimates is evaluated by computing standard errors, which increase with the level of disaggregation. In general:

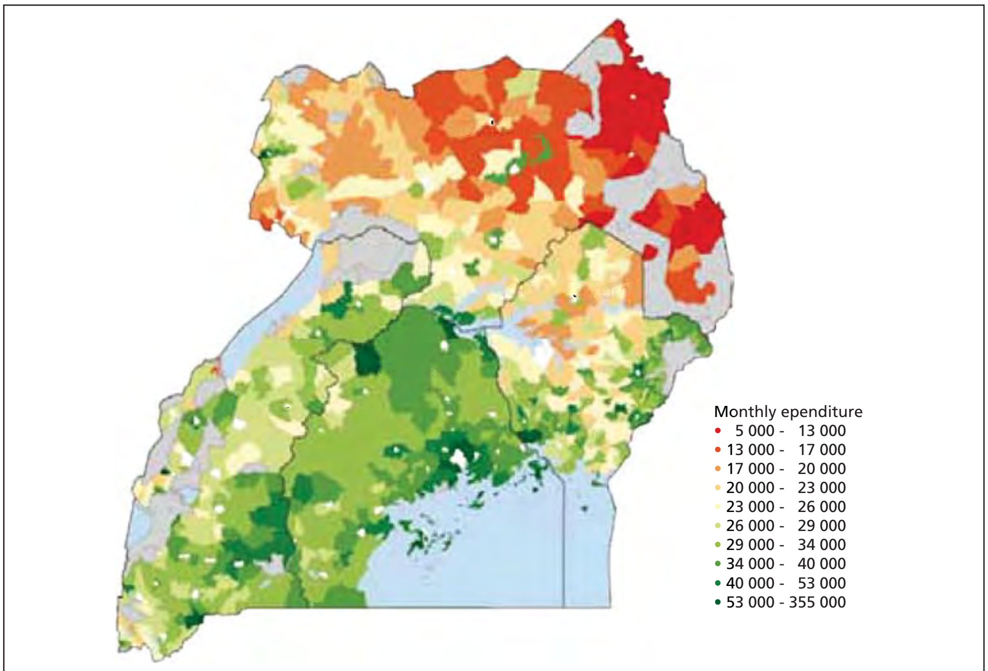
$$y_i = A_i'\beta_i + \varepsilon_i$$

where y_i is the welfare indicator for household *i*, A_i' is a vector of independent variables (and associated parameters, β_i) common to the welfare survey and the census and ε_i is a normally distributed error term.

Small area poverty estimates have been made for a number of countries, for example Ecuador (Hentschel *et al.* 2000), South Africa (Alderman *et al.* 2000; Statistics South Africa 2000), Nicaragua (Arcia *et al.* 1996); Vietnam (Minot *et al.* 2003); Epprecht and Heinimann 2004); Kenya (Ndeng'e *et al.* 2003); and Uganda (Emwanu *et al.* 2003; 2007).

At the time that Rogers *et al.* (2006) published their working paper, small area estimates (SAE) of welfare had not been produced for the UNHS-2 household survey data, so direct comparisons with the environmental approach were not possible. Since then, however, SAE poverty mapping has been applied to the same household survey used in the present analysis. Emwanu *et al.* (2007) combined information from the 2002/03 UNHS-2 (UBOS 2003) and the 2002 Population and Housing Census (UBOS 2002) to develop poverty maps at district, county and sub-county levels. The sub-county estimates are shown in Figure 2.

Figure 2. Small area (sub-county) estimates of average rural monthly adult equivalent expenditure (in Uganda shillings).



Source: Emwanu *et al.* (2007).

ENVIRONMENTAL TIME SERIES DATA

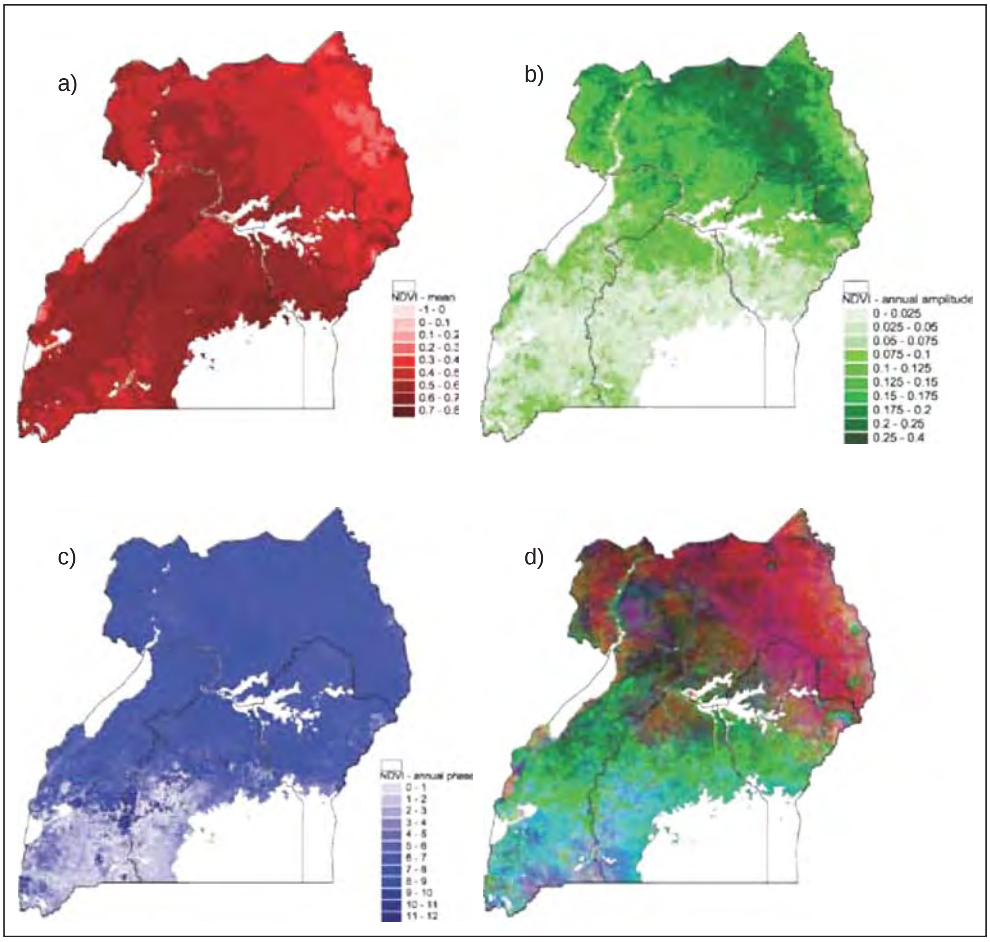
The majority of the explanatory variables used in the regression modelling were satellite-derived, and most came from the 1km global Advanced Very High Resolution Radiometer (AVHRR) dataset made available by the National Aeronautics and Space Administration (NASA) Pathfinder program. These data were processed by the Pathfinder program only for a limited number of months between 1992 and 1996. The data were aggregated into synoptic monthly (maximum value) composites to give a record of monthly changes in an average year. One synoptic series was produced for each of the following: the middle infra-red (MIR) - AVHRR channel 3; Land Surface Temperature (LST) - produced by combining information from AVHRR channels 4 and 5); the Normalised Difference Vegetation Index (NDVI) – produced by combining information from AVHRR channels 1 and 2; air temperature (Tair) - produced by combining LST with the (NDVI); and Vapour Pressure Deficit (VPD) - a combination of satellite and ground-based meteorological data. In addition to the five products derived from AVHRR data, information from the European geostationary Meteosat satellite in the form of a rainfall surrogate, the Cold Cloud Duration (CCD), was obtained from the FAO ARTEMIS program[1].

The original AVHRR imagery is in the Goode's Interrupted Homolosine projection, and the CCD imagery in the Hammer-Aitoff projection (a variant of the Lambert projection). Each data series was temporally Fourier-processed to produce 10 separate data layers; the mean (1 layer), the phases and amplitudes of the annual, bi-annual and tri-annual cycles of change (6 layers in all), the maximum, minimum (2 layers) and the variance (*i.e.* original channel variance, not that of the Fourier series) (1 layer). The Trypanosomiasis And Land use in Africa (TALA) Research Group in Oxford, UK, has developed this unique way of processing multi-temporal satellite data that captures the seasonality of natural habitats and is thus ideal for describing biological processes that depend on them. Temporal Fourier processing has all the statistical advantages of any good ordination technique applied to satellite data (Fourier variables are statistically independent of each other), and the additional advantage that the condensed outputs may be interpreted in a biological context. Further details of temporal Fourier analysis of satellite data are given in Rogers *et al.* (1994); Rogers *et al.* (1996); Rogers (1997); and Rogers (2000).

After temporal Fourier processing, the data were re-projected to the longitude/latitude system by bi-linear interpolation to a nominal pixel resolution of 0.01 degrees (about 1.1 km at the equator). For those data layers at an original spatial resolution coarser than 1km (hence also of 0.01 degrees), the data were interpolated to the same spatial resolution: this applied to the VPD and CCD imagery. A far more thorough account of how the environmental data used in this analysis were produced is provided in Rogers *et al.* (2006), where examples of imagery can also be found. Figure 3 shows, as an example, Fourier processed imagery of the mean annual, annual amplitude and annual phase for NDVI.

[1] METEOSAT data were kindly provided by Fred Snijders of the ARTEMIS program, FAO.

Figure 3. Fourier processed NDVI imagery for: a) mean annual, b) annual amplitude, c) annual phase, and d) a 3 band false-colour composite of these layers, which summarises the spatial variability in NDVI.



LIVESTOCK PRODUCTION SYSTEMS MAP

Seré and Steinfeld (FAO 1996) developed a classification of livestock production systems based on agro-ecology and the distinction between mixed and pastoral, irrigated and rain-fed, and urban/landless areas. Arising from this is one of the more widely used classifications of livestock production systems, developed and mapped by the International Livestock Research Institute (ILRI) Thornton *et al.* (2002).

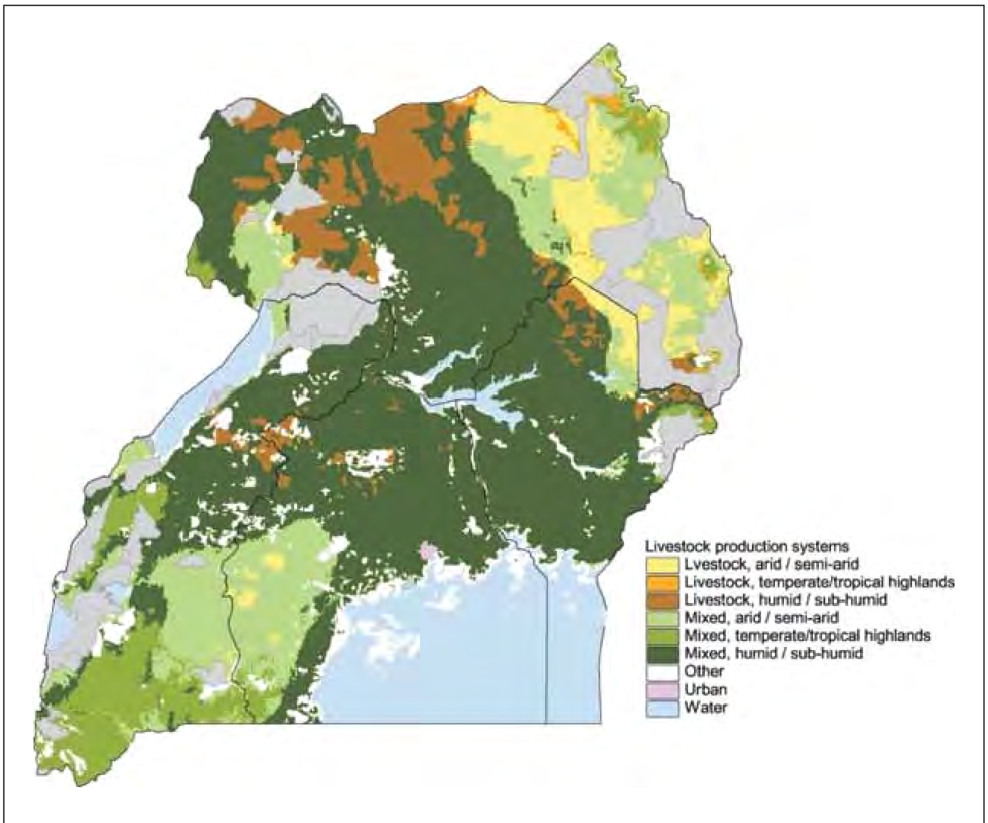
The classification is based on four modes of production: livestock grazing; rain-fed mixed crop and livestock production; irrigated mixed crop and livestock production; and landless livestock production. These are further split among three agro-ecological zones defined by LGP and temperature: arid and semi-arid; humid and sub-humid; and temperate or tropical highlands. Data on land cover, irrigation, LGP, temperature, elevation and population density were incorporated into the original classification, as described in detail in Thornton *et al.* (2002) and in Kruska *et al.* (2003). This classification has been used to stratify many analyses (some described in FAO 2007) and, having climatic and population variables as input data, has enabled the classification to be re-evaluated in response to different scenarios of climate and population change (Thornton *et al.* 2008).

The classification of Thornton *et al.* (2002) was originally produced for the developing world, but it has recently been extended globally, using essentially the same methods and more recent and detailed data (Robinson *et al.* 2011). Figure 4 shows version 4 of the mapped livestock production systems for Uganda in which the mixed irrigated classes have been merged with the mixed rain-fed classes in similar agro-ecological zones – irrigation being relatively unimportant in Uganda.

OTHER SPATIAL DATA

In addition to the above, the following layers were considered as potential predictor variables for regression modelling: slope and elevation from the CSI-SRTM void filled 90 m Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) data (v4.1, Reuter *et al.* 2007); human population density circa year 2000 from the Global Rural Urban Mapping Project (GRUMP) dataset (CIESIN 2004); access to markets as measured in travel time in hours to the nearest populated centre (Pozzi *et al.* 2009); and cattle, sheep, goat and pig densities[2] (FAO 2007). All of these additional data layers were converted to the same geographic reference system as the satellite data for Uganda, and similarly aggregated, by averaging, for the models at different spatial resolutions. Maps of the variables that were selected for use in the regression modelling are shown in Figure 8.

Figure 4. Summary of livestock production systems in Uganda.



Source: Robinson *et al.* (2011).

[2] Poultry densities were not included since their distribution closely follows that of the human population.

Methods

COUNTRY-WIDE OLS REGRESSION

In the OLS multiple regression model, the dependent variable y (the chosen measure of poverty, here household expenditure) is statistically related to a set of N independent variables x as follows,

$$y_i = \beta_o + \sum_{j=1}^{N} x_j \beta_j + \varepsilon_i \quad \text{where } i = 1 \text{ to } M$$

where i is an index of the number of points (M) for which data are available, β_0 is the intercept, β_j are the beta-coefficients for each dependent variable, and ε is a randomly distributed error term. The reader is referred to standard texts on regression such as Draper and Smith (1988) for a fuller explanation of OLR and, for specific application to geographical problems, to Griffith and Amrhein (1997)

In addition to the beta-coefficients, the following statistics were calculated in order to evaluate model accuracy:

- The residuals at each location, and the residual sum of squares.
- The standard errors associated with each β term.
- A t statistic for each independent variable.
- A coefficient of determination statistic (R^2).

A single regression model was fitted to the aggregated household data and the environmental variables at those same locations across Uganda. The β coefficients were then applied to the predictor variables to extrapolate the relationships to estimate the average rural monthly per adult equivalent household expenditure in all pixels.

A series of diagnostic tests was also carried out to ensure that the model met the following assumptions, required for linear regression.

- Homoscedasticity – the variance of the error term must be constant for each value of y. To check this the residuals were plotted against y. Ideally, there should be no obvious pattern.
- No multicollinearity – no strong correlation should be observed among the independent variables. Bivariate collinearity was checked for with scatter plots and correlations between each pair of independent variables, and was assessed with a variance inflation factor test.
- Linearity – there should be a linear relationship between each independent variable and the dependent variable. This can be assessed with a scatterplot matrix for all variables. Non-linearity does not invalidate the OLS model but it does mean that the beta coefficients cannot fully capture the relationship. The dependent variable was transformed to ensure linearity, but the independent variables were not.
- Independence of error terms – successive residuals should not be correlated. A Durban-Watson statistic was used to check for such autocorrelation.

REGIONAL OLS

A single regression model is unlikely to capture the relationships between expenditure and the environment across an entire country, because the influence of each environmental factor, and the complex interactions among them, are likely to vary from location to location. Instead, a series of local or regional regression models might be more appropriate. Because of the potential importance of livestock, various aggregations of the livestock production system map were used to partition Uganda into zones for which to derive separate regression models. Only one of the six zones present in Uganda contained enough sample households to be treated independently, so the following three sets of aggregated categories were chosen:

- Three 'climate zones': (i) arid and semi-arid, (ii) humid and sub-humid and (iii) temperate or tropical highlands.
- Two 'farming systems': (i) livestock only and (ii) mixed crop and livestock.
- One 'dominant system': the mixed, humid and sub-humid system – the largest in Uganda.

The OLS coefficients from the first two can be combined to create country wide maps of predicted average rural monthly adult equivalent expenditure in all pixels. The 'dominant system' regression coefficients can only be extrapolated to pixels within that system. It is important to stress that within each of the bulleted headings above the categories are mutually exclusive (e.g. an area can be either 'livestock only' or 'mixed crop and livestock' only; it cannot be both, or a mixture of both). However there is considerable overlap between the bulleted headings since they form alternative ways of zoning what are frequently the same areas. Thus the regional OLS analyses using the three categorisations given above overlap considerably in the data points used (see later).

GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

An OLS regression model can be converted into a Geographically Weighted model by substituting each beta coefficient (the intercept and the dependent variable coefficients) with its local counterpart, such that the beta-coefficients can vary across space:

$$y_i = \beta_{o(u_i,v_i)} + \sum_{j=1}^{N} x_j \, \beta_{j(u_i,v_i)} + \varepsilon_i$$

where $i = 1$ to M, *and* (u_i,v_i) is the location in geographic space of the *i*th observation. A set of beta-coefficients (and hence a regression model) is estimated at each location based only on neighbouring, geographically weighted data points. Normally variables from data points farther away from the point in question (where y_i was measured) have lower weights than points closer to it and so contribute less to the regression results. If, however, $\beta(u_i,v_i)$ is constant for all (u_i,v_i) then the OLS model holds, i.e. OLS is a special case of GWR. The geographically weighted local counterparts of the residuals, standard errors, t and R^2 values (and any other associated statistic) can also be generated at each location.

Fotheringham *et al.* (2002) have developed GWR into a comprehensive statistical method. A key feature is the ability to calibrate the spatial weighting function to identify the bandwidth, i.e. the number of, or proximity of neighbouring points included, that results in a 'best-fit' model.

The estimated beta-coefficients at each location are dependent on the bandwidth and type of kernel (or weighting scheme) that is used in the model. Here, the bi-square kernel was used, defined as:

$$W_i = \begin{cases} \left(1 - \left(\frac{D_i}{h}\right)^2\right)^2 f \quad D_i < h \\ 0 \;\; f \quad D_a \geq h \end{cases}$$

Where W_i is the weight assigned to point i (from 0 to 1), h is the bandwidth and D_i is the distance from the centre of the kernel to point i.

The most appropriate bandwidth can be chosen by means of a cross-validation (CV) procedure where the model is run for a range of bandwidths and a least squares criterion is applied to find the bandwidth that minimises the sum of the squared errors between y and the estimated value of y (y'). The equation below states that for bandwidth h, y'_i is computed whilst omitting the data from point i. This omission of the central point means that when h is very small, the model is calibrated only on its neighbouring points and not on itself. If point i were not omitted the CV score would tend to zero as h tends to zero, (and hence the weighting for all points except i becomes negligible) and so y'_i would tend to y_i, at each location.

$$CV = \sum_{i=1}^{N} \left(y_i - y'_i(h) \right)^2$$

The bandwidth can be defined in map units or as the number of data points (nearest neighbours) to include at each location. The use of map units, whilst ideal, can only be justified if the data points are evenly distributed over the study area. The survey households are not, so the bandwidth was determined in terms of number of nearest f neighbours.

Once the model has been calibrated and the best bandwidth identified, the GWR is re-run using the best bandwidth, and a series of computationally intensive tests is run to evaluate significance in spatial variation among the GWR parameters. GWR produces localised versions of the OLS regression outputs, so in place of a table of results summarising the beta coefficients, t values, standard errors etc., localised versions of these outputs are produced for each household or household pixel, and these can be mapped. The local R^2 and t values can be interpolated to give a visual representation of the goodness of fit of the model and to map areas where the coefficients are significant, but interpretation of these local statistics is not as straightforward as it is with their OLS equivalents. Furthermore, the GWR model can be applied to the remaining rural pixels in Uganda to create an estimated rural monthly adult equivalent expenditure map. The beta coefficients can also be mapped, to show the spatially varying relationships (non-stationarity) between poverty and the environmental variables included. Finally, these beta coefficient maps can be viewed as multiband images, or clustered, in order to identify regions with common spatial relationships between poverty and environmental variables.

WORKFLOW

The dependent variable - rural monthly adult equivalent expenditure – was first aggregated to match the finest resolution of the environmental variables, 0.01 degrees (approximately 1.1 km at the equator). These values were then transformed using a Box-Cox transform (Box and Cox 1964), resulting in a normally distributed dependent variable. This process was then repeated at a series of spatial resolutions matching those of the environmental data (broadly successive doubling of pixel dimensions, quadrupling their size).

A model was built, using the 0.01 degree data, relating expenditure to environmental conditions based on previous poverty mapping work and correlation analysis. The same variables were then used at all other (coarser) spatial resolutions, and any changes in the relative importance, sign and significance of the independent variables noted.

The above approach was applied to each of the methods used, OLS, regional OLS and GWR, providing regression results and maps of predicted average rural monthly adult equivalent expenditure. A bootstrap procedure was used to compute four goodness of fit metrics and their standard errors: a) Root Mean Square Error (RMSE); b) Mean Absolute Error (MAE); c) Mean Absolute Percentage Error (MAPE); and d) the R^2 of the observed versus the expected expenditure for each regression model (OLS, regional OLS and GWR) at all resolutions. These statistics helped to determine the resolution that provided the best trade-off between predictive accuracy and spatial precision. The same four metrics were further computed for the SAE expenditure maps at district, country and sub-country levels.

Finally, the GWR coefficients were mapped at the best resolution and the spatial variation in the coefficients was investigated.

All analyses were carried out in 'R' (V2.9.2)[3] (R-Development-Core-Team 2009) running on a 32bit version of Windows XP (SP3). The following R libraries were used:

- MASS (V7.2-48) - Functions and datasets to support Venables and Ripley, 'Modern Applied Statistics with S'.
- car (V1.2-15) – Companion to Applied Regression.
- relaimpo (V2.1-2) - Relative importance of regressors in linear models – Non US version[4].
- qpcR (V1.2-1) – Used for computing Akaike's Information Criterion.
- psych (V1.0-78) – Used for basic statistical summaries.
- gvlma (V1.0) - Global Validation of Linear Models Assumptions.
- spgwr (V0.6-2) - Geographically weighted regression.
- boot (V1.2-39) – Bootstrapping regression models to generate confidence limits and standard errors.

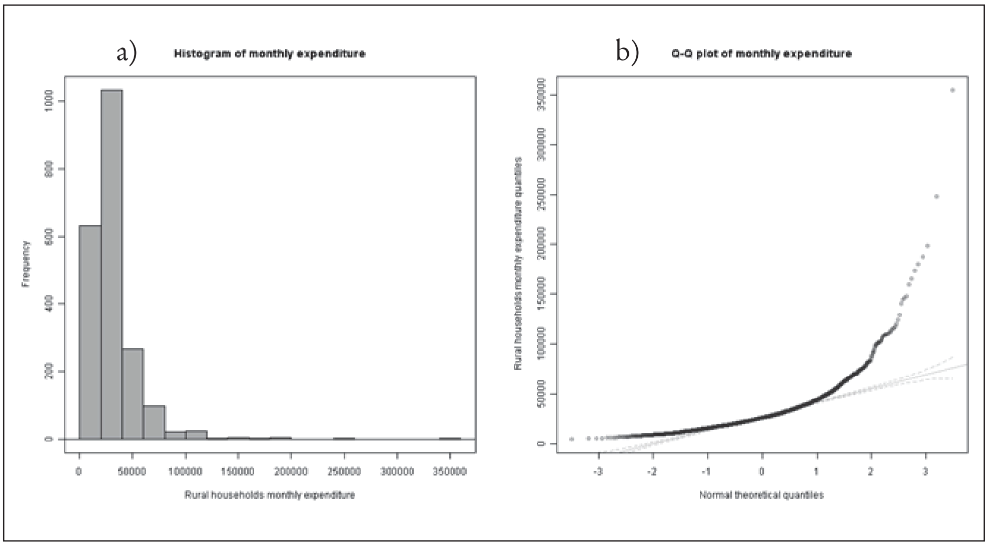
Libraries or packages and their dependencies can be installed and updated in R, on the command line. The R code used is available from the authors.

[3] http://cran.r-project.org and http://www.r-project.org
[4] http://prof.beuth-hochschule.de/groemping/relaimpo

TRANSFORMATION OF THE EXPENDITURE DATA

A simple histogram and q-q plot[5] of the aggregated rural monthly adult equivalent expenditure values confirmed their distribution to be far from normal (Figure 5).

Figure 5 Distribution of the rural monthly adult equivalent expenditure at 0.01 degree resolution (a), and q-q plot of quantiles against a theoretical, normal distribution (b).



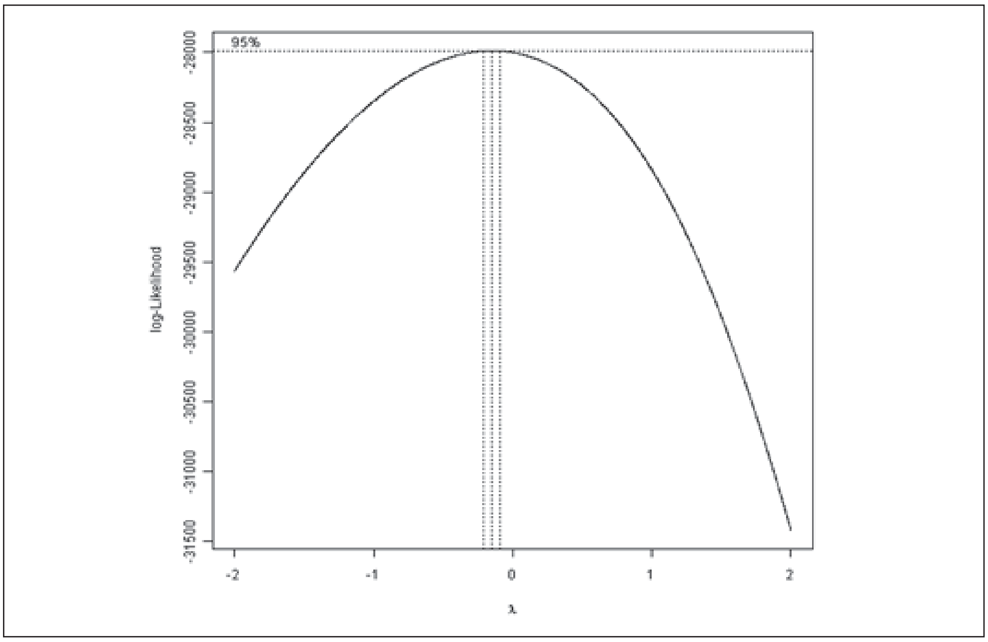
The Box-Cox power transform (Box and Cox 1964) was used to normalise the distribution, which uses a power parameter, λ, and has the general form:

$$y^{bct} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & f \quad \lambda \neq 0 \\ \ln(y) & f \quad \lambda = 0 \end{cases}$$

A simple procedure in R computes and plots log-likelihoods for λ with 95 percent confidence limits (Figure 6). If zero lies within the confidence limits then ln(y) would be the more appropriate transform. In this case, λ was -0.151 and zero was not contained within its confidence limits, suggesting the Box-Cox transform to be the more appropriate transformation.

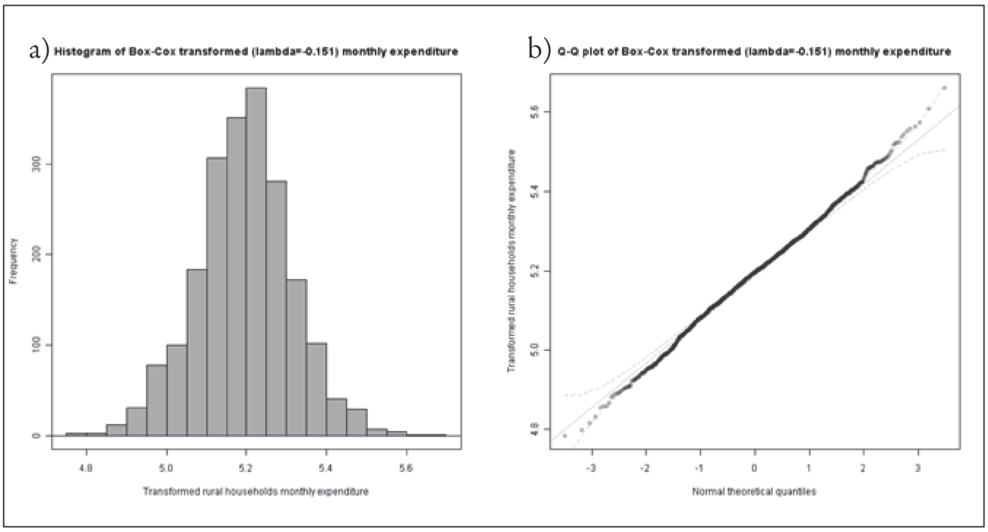
5 a Q-Q plot ('Q' stands for quantile) is a probability plot, a kind of graphical method for comparing two probability distributions, by plotting their quantiles against each other. Here we compare the household data distribution against a normal distribution.

Figure 6. Log-likelihood for **λ** in the Box-Cox transformation at 0.01 degrees.



The histogram and q-q plot (Figure 7) of these transformed expenditure data show the distribution to be much more normal.

Figure 7. Distribution of the transformed rural monthly adult equivalent expenditure at 0.01 degree resolution (a), and q-q plot of quantiles of transformed data against a theoretical, normal distribution (b).



The inverse transform was applied to obtain the resulting predicted welfare in the original units:

$$y = \begin{cases} (\lambda y^{bct} + 1)^{1/\lambda} & f \quad \lambda \neq 0 \\ e^{y^{bct}} & f \quad \lambda = 0 \end{cases}$$

Table 2 shows the λ values for each resolution along with the number of pixels that contained household data. At resolutions coarser than 0.75 degrees there are too few pixels containing data to fit reliable regression models.

Table 2. Pixel counts and λ values for the Box-Cox transformation at each spatial resolution (in decimal degrees).

Cell size	0.01	0.02	0.03	0.05	0.10	0.15	0.20
# of pixels	2 088	1 279	1 086	813	539	399	280
λ	-0.151	-0.074	-0.027	0.035	0.104	0.206	0.300
Cell size	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**	**0.75**
# of pixels	206	167	120	103	82	75	36
λ	0.359	0.325	0.250	0.472	0.113	0.465	0.898

VARIABLE SELECTION AND DESCRIPTION

Based on the outcomes of previous studies, preliminary analyses and data availability the following were selected as independent, predictor variables: (i) mean annual NDVI (ndvi); (ii) mean annual Vapour Pressure Deficit (vpd); (iii) slope (slp); (iv) goat density (goat); (v) cattle density (cattle); (vi) travel time to markets (dist); and (vii) population density (grump).

These seven variables, at 0.01 degrees resolution, are shown in Figure 8. NDVI and VPD show variation in climate from the more humid central and southern regions to the arid northern and eastern regions. Goat densities are highest in the northeastern and southwest regions whereas cattle are found in a broad band spanning from the southwest to the northeast; the so-called 'cattle coridoor'. Population density (grump) is higher, and access to markets (dist) better, in the central and southern regions than elsewhere. Finally, slp reflects well the mountainous terrain in the eastern and southern regions.

Table 3 shows a correlation matrix of the dependent (ybct) and independent variables. Firstly this demonstrates that there are no major collinearities among the independent variables. Secondly it shows that two of the independent variables, NDVI (+ve) and VPD (-ve), have stronger correlations with ybct than the other variables, goat (-ve), cattle (+ve), slp (+ve), grump (+ve), dist (-ve). The signs are broadly as expected, with the exception of slp, although the correlation is very weak, and possibly goat density (goats are predominant in the less wealthy pastoral and agro-pastoral areas of the northeast of Uganda). Table 4 gives a numerical summary of each variable, showing the skewed distributions of several of them.

In conclusion, these variables show some degree of correlation with per-adult equivalent expenditure, have little collinearity, and, in general have the expected sign.

Figure 8. Independent variables used in the regression models.

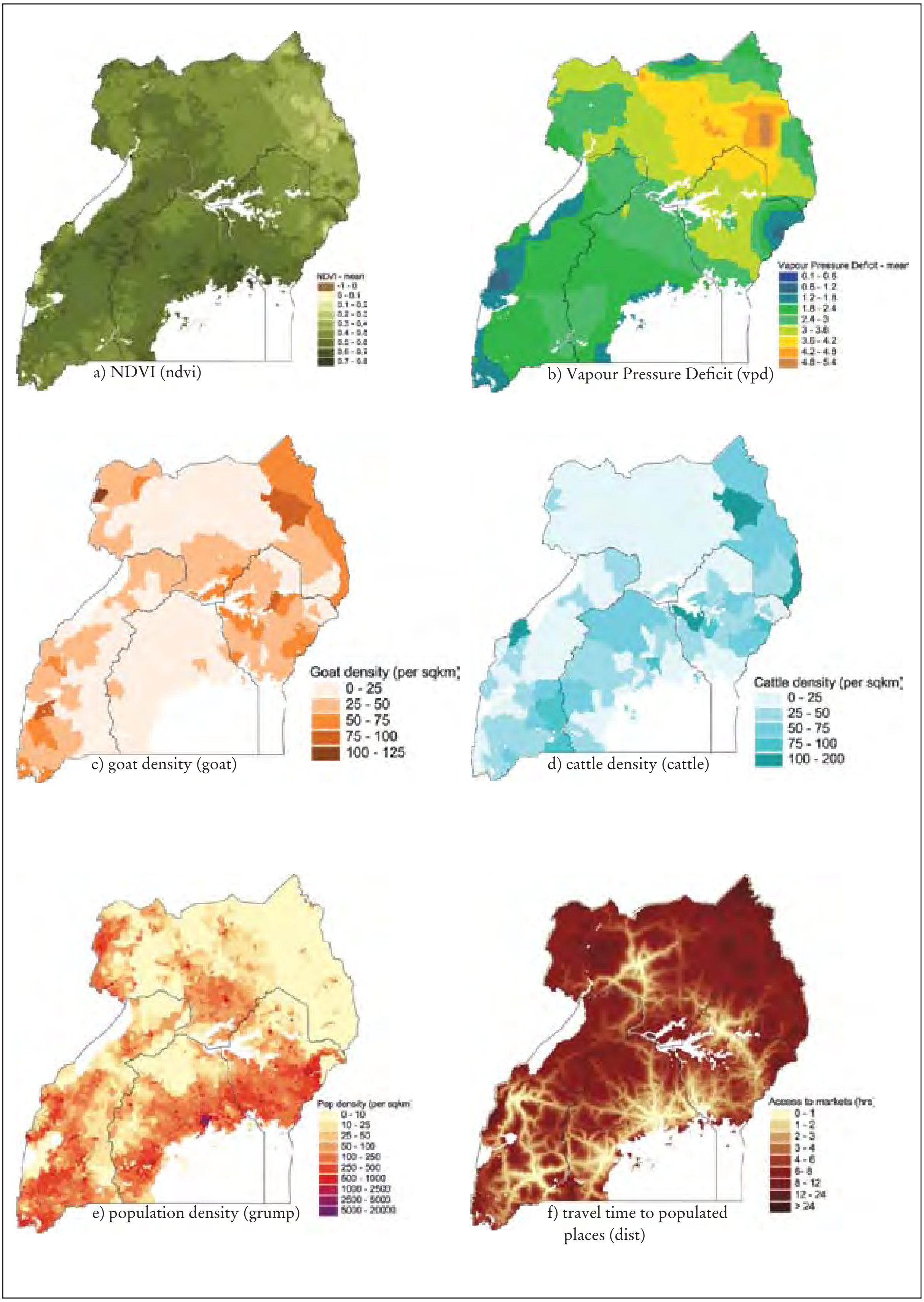
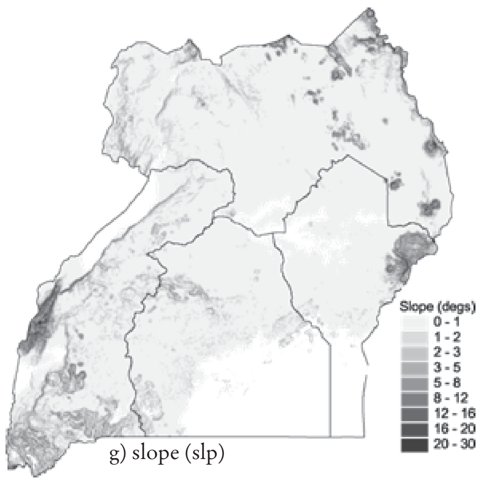


Figure 8 (cont).



g) slope (slp)

Table 3. Correlation matrix for the dependent (ybct) and independent variables.

	ndvi	vpd	goat	cattle	slp	grump	dist
ybct	0.307	-0.297	-0.178	0.045	0.041	0.194	-0.241
ndvi		-0.382	-0.254	-0.143	-0.017	0.072	-0.208
vpd			0.113	-0.092	-0.509	-0.322	0.094
goat				0.273	0.096	0.13	0.176
cattle					0.031	0.084	-0.037
Slp						0.265	0.038
grump							-0.319

Table 4. Descriptive statistics for the dependent (ybct) and independent variables at 0.01 degrees resolution.

Momental statistics (n=2 088)					
Variable	**Mean**	**Std Err Mean**	**Std Dev**	**Skewness**	**Kurtosis**
ybct	5.19	0.003	0.12	-0.01	0.34
ndvi	0.52	0.001	0.07	-2.23	18.57
vpd	2.62	0.015	0.68	0.08	-0.44
goat	32.43	0.427	19.50	0.95	1.90
cattle	33.37	0.471	21.54	1.43	2.47
slp	1.13	0.032	1.48	3.02	10.92
grump	190.62	4.415	201.72	4.47	37.87
dist	254.12	3.276	149.71	1.23	3.65

OLS RESULTS

In this section the full set of regression results and diagnostics are presented for the 0.01 degree resolution analysis, and summaries for the analyses conducted at coarser resolutions.

The results show each independent variable to be significant at the 1 percent level or better (Table 5). The overall model is significant, explaining 18.52 percent (the multiple r-squared value in Table 5, expressed as a percentage) of the variability in the rural monthly adult equivalent expenditure at 0.01 degrees resolution. Throughout, R^2 rather than adjusted R^2 has been used as it is a direct measure of the model's ability to explain the variance in the data (adjusted R^2 values remove the effect of collinearity of the predictors, which was slight here).

Table 5. Descriptive statistics for the dependent (ybct) and independent variables at 0.01 degrees resolution (c. 1.1 km at the equator).

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	Signifi.[1]
(Intercept)	5.155e+00	3.246e-02	158 829 .	< 2e-16	***
ndvi	3.085e-01	4.084e-02	7.555	6.23e-14	***
vpd	-3.521e-02	4.737e-03	-7.432	1.55e-13	***
goat	-6.884e-04	1.338e-04	-5.143	2.95e-07	***
cattle	3.883e-04	1.160e-04	3.347	0.000832	***
slp	-5.789e-03	1.947e-03	-2.974	0.002974	**
grump	6.071e-05	1.337e-05	4.539	5.97e-06	***
dist	-1.003e-04	1.733e-05	-5.789	8.16e-09	***

Residual standard error: 0.1067 on 2 080 degrees of freedom; multiple R-squared: 0.1852; adjusted R-squared: 0.1825; F-statistic: 67.55 on 7 and 2 080 DF; p-value: < 2.2e-16; AICc: -3 410.745.

[1] *** p<0.001 ; ** p<0.01 ; * p<0.05

The signs of the coefficients were as expected: NDVI, cattle density and population density all had a positive influence on rural monthly adult equivalent expenditure, while VPD, goat density, slope and travel time to markets all had a negative influence.

A set of standard diagnostic plots is shown in Figure 9. The graphs in Figures 9a and 9b plot the residuals against the fitted or predicted values. Ideally the variance in the residuals should be constant regardless of the predicted value and no obvious pattern in the scatter should be evident. Evidence of non-constant variance (heteroskedasticity) would appear as a fan-shaped pattern of increasing variance. Figure 9c is a q-q plot confirming that the residuals are normally distributed. The final graph, Figure 9d, measures the 'leverage', which refers to the degree to which some points unduly influence the regression model – high leverage means high influence.

Two tests were performed to see whether key assumptions of the OLS model were being violated. A formal chi-square test for non constant variance gave a value of 4.279254 (df = 1), with p = 0.03858027; any value of $p < 0.05$ confirms the unchanging nature of the residuals (Figures 9a and 9b). Multicollinearity in the independent variables was further tested for, using Variance Inflation Factors (VIF), which are indices that measure how much the variance of a coefficient increases because of collinearity. The smaller these VIF scores the better, with values of less than 2.0 being acceptable (Table 6). Both tests therefore confirmed that the OLS model did not violate the assumptions of OLS regression.

Figure 9. Diagnostic plots of the 0.01 degree OLS regression.

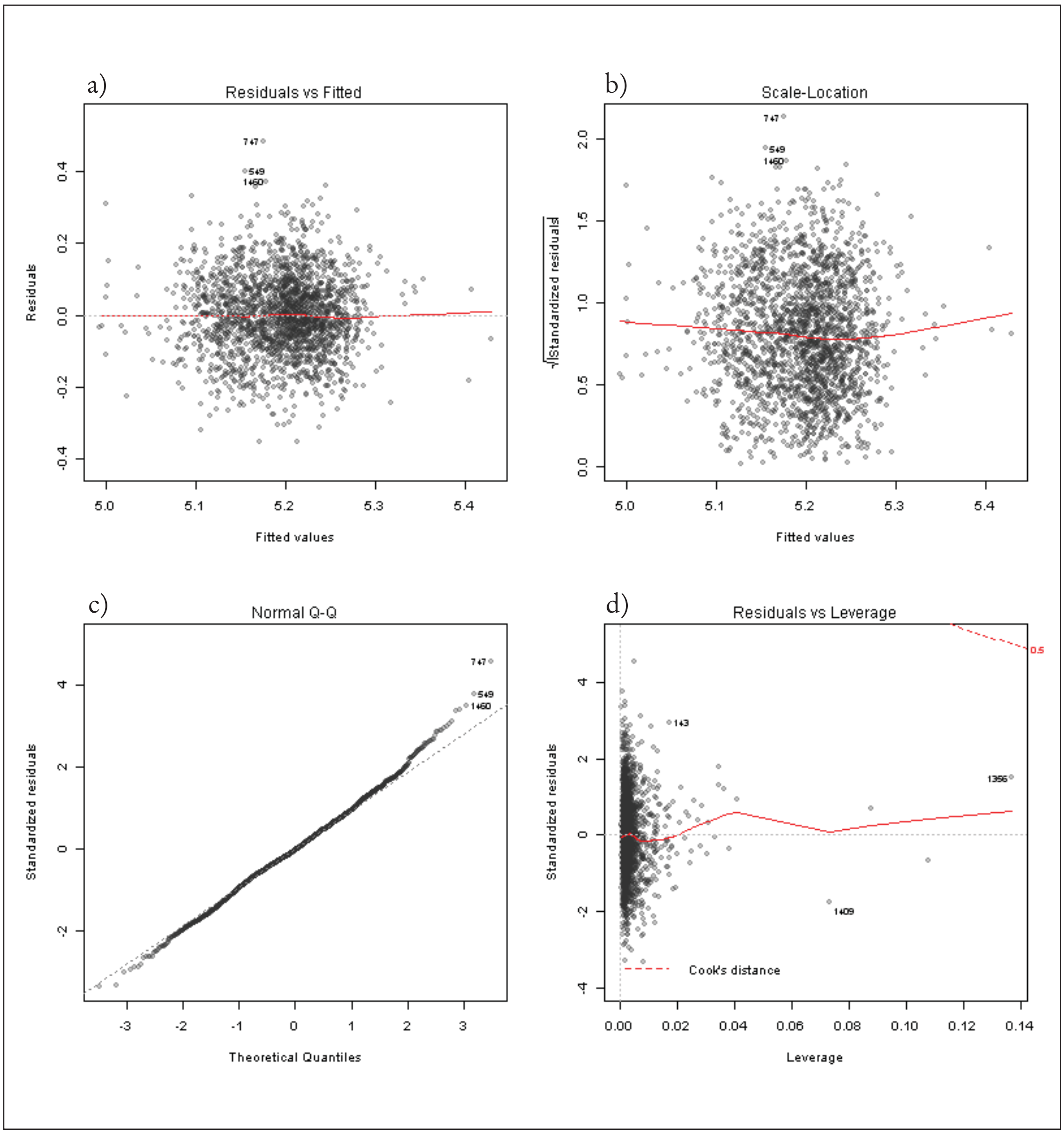


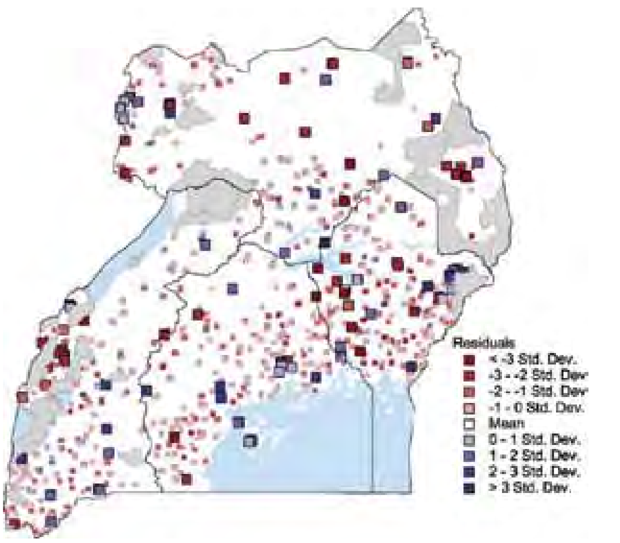
Table 6. Results of a Variance Inflation Factor (VIF) test for multicollinearity.

Variable	ndvi	vpd	goat	cattle	slp	grump	dist
VIF	1.174	1.374	1.118	1.070	1.230	1.156	1.111

Note: VIF values less than 2 indicate collinearity not to be a problem.

Figure 10 maps the standardized residuals and highlights the 110 points with high leverage as defined by the conventional cut-off of a Cook's distance greater than 4/n, where n is the number of observations (Bollen and Jackman 1990). Since these points were not spatially clustered, but rather distributed evenly across Uganda, there was no good theoretical reason to remove them.

Figure 10. Standardized residuals for the 0.01 degree OLS regression with high leverage points highlighted with larger symbols.



The relaimpo library in R was used to conduct a Lindeman, Merenda and Gold (LMG) analysis (Linderman *et al.* 1980), that quantifies the relative importance of each predictor variable in determining the model's explanatory power . The LMG analysis "computes the sequential sums of squares from the linear model…for an overall assessment by averaging over all orderings of regressors" (Grömping 2007) resulting in a decomposition of R^2 by variable. The resulting plot is shown in Figure 11. NDVI and VPD are the two most important variables at this scale for the countrywide regression, with travel time to populated centres third. These three variables accounted for almost 80 percent of the explanatory power of the model.

Finally, the coefficients from the OLS model were used to predict average rural monthly adult equivalent expenditure at 0.01 degrees resolution across Uganda. This map (which does not necessarily represent the most appropriate model or resolution for estimating the expenditure) is shown in Figure 12.

Figure 11. Estimate of the relative importance of the independent variables for the 0.01 degree OLS regression, including 95 percent confidence limits.

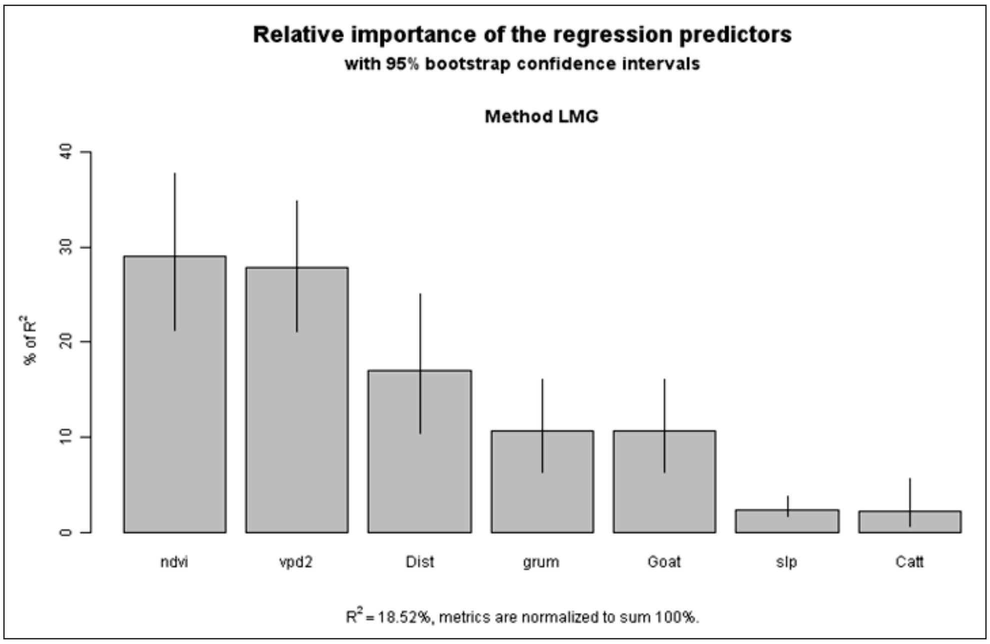
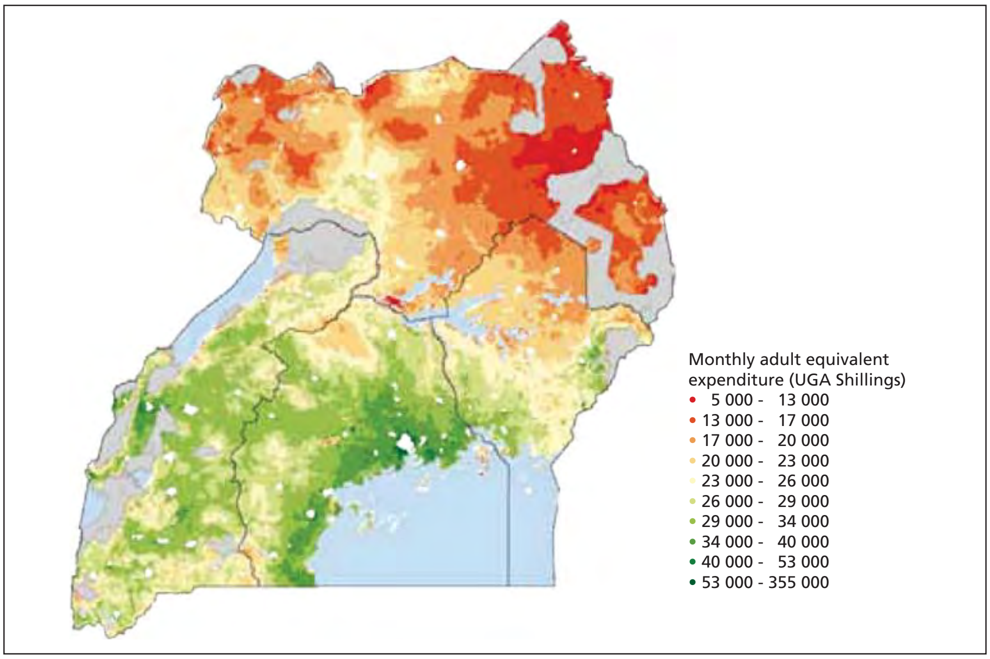


Figure 12. Predicted average rural monthly adult equivalent expenditure.



Note: based on a country-wide OLS regression model at 0.01 degrees resolution (c. 1.1 km at the equator).

The map bears a strong resemblance to the SAE map of rural monthly adult equivalent expenditure (Figure 2), with lower expenditures in the northern and especially eastern regions and higher expenditures in the southern and central regions, especially around Kampala and Lake Victoria.

The same analysis was performed at each spatial resolution, for a subset of which the OLS model outputs are given in Table 7. As a rule of thumb there should be at least 50 independent data points for each independent variable, so at least 350 data points are required in this case (7 independent variables). Any OLS results at 0.20 degrees and coarser resolutions, therefore, should be treated with caution. Focussing on the results from cell sizes of 0.01 to 0.15, Table 7 shows that R^2 tends to increase, while the sign and relative importance of the variables was more or less constant: NDVI, VPD, population density and goat density were the most important variables, while travel time to markets, slope and cattle density were generally less important.

Table 7. OLS model summary for all resolutions.

Model results			Significance[1], sign (+/-) and relative importance (1-7)						
Cell size	Points	R^2	ndvi	vpd	goat	cattle	slp	grump	dist
0.01	2 088	0.185	*** + 1	*** - 2	*** - 5	*** + 7	*** - 6	*** + 4	*** - 3
0.02	1 279	0.176	*** + 4	*** - 1	*** - 2	** + 6	*** - 5	*** + 3	*** - 5
0.03	1 086	0.206	*** + 3	*** - 1	*** - 2	** + 7	*** - 6	*** + 4	*** - 5
0.05	813	0.227	*** + 4	*** - 1	*** - 3	** + 7	*** - 5	*** + 2	- 6
0.10	539	0.292	*** + 4	*** - 1	*** - 3	** + 7	*** - 6	*** + 2	- 5
0.15	399	0.290	* + 4	*** - 1	*** - 3	* + 7	*** - 5	*** + 2	- 6
_0.20_	_280_	_0.364_	- 6	*** - 1	*** - 3	+ 7	*** - 4	*** + 2	- 5
_0.25_	_206_	_0.371_	- 6	*** - 1	** - 3	+ 7	*** - 4	*** + 2	- 5
_0.30_	_167_	_0.421_	- 7	*** - 1	*** - 3	+ 5	*** - 4	*** + 2	- 6
_0.35_	_120_	_0.513_	- 6	*** - 1	*** - 3	** + 7	*** - 4	*** + 2	- 5
_0.40_	_103_	_0.409_	+ 4	*** - 1	** - 3	+ 7	* - 6	** + 2	+ 5
_0.45_	_82_	_0.587_	* - 7	*** - 1	** - 3	*** + 6	*** - 2	** + 4	- 5
_0.50_	_75_	_0.527_	+ 6	*** - 1	- 3	+ 7	* - 4	* + 2	- 5
_0.75_	_36_	_0.614_	- 5	*** - 1	- 4	+ 7	*** - 3	+ 2	- 6

[1] *** p<0.001 ; ** p<0.01 ; * p<0.05

Note: Most important variables in bold. Table rows in italics are for regressions with fewer than the minimum recommended number of data points (see text).

REGIONAL OLS RESULTS

Analyses of the six regional sub-sets of the data were carried out in the same way as described above for the single country-wide regression. Not all results are presented here since the regional models encountered the same problem of insufficient data points as did the country-wide model at coarser spatial resolutions. The main purpose here was to determine whether there were differences in the sign and significance of the regression coefficients, and of the relative importance of the regression variables, in the different zonations used. Table 8 summarises these three features for the six regions at 0.01, 0.05 and 0.10 degrees spatial resolution.

It follows from the description earlier of the zonation schemes used here that there is considerable overlap between the ones containing large numbers of data points; in particular the humid and sub-humid climate zone; the mixed crop and livestock systems; and the intersection of these - the dominant mixed, humid and sub-humid system (Figure 4). It is therefore perhaps not surprising that the regional OLS results are similar to each other for these zones, and in some respects to the OLS country-wide regression. The signs of the coefficients were consistent across the regions and across resolutions; VPD and NDVI were the two most important variables while cattle and slope were consistently the least important. VPD was nearly always the most important variable in the OLS country-wide model, but the 2^{nd} and 3^{rd} placed variables varied by region and resolution.

Three of the regional analyses (the climate zones: arid and semi-arid, and temperate/tropical highland; and the farming system: livestock-only) lacked data points at all three resolutions.

As in the OLS results, the R^2 values generally increased as the cell size increased and the number of data points decreased. This perceived improvement in the model at larger cell sizes needs to be treated with caution but may be important. Random data showing no relationship between environmental variables and household expenditure would not show any improvement in r-squared values if they were progressively combined by averaging, as here, whereas any real relationship between these two variables is likely to be quite noisy at the finest spatial resolution and less noisy at aggregated resolutions, the effect of aggregation being to cancel out noise, and hence to reveal more of the true 'signal'

The predicted expenditure for all six regions was computed and then combined to create three maps of expenditure at 0.01 degrees resolution as follows: (i) climate – a combination of models for the arid and semi-arid, temperate and tropical highlands and humid and sub-humid regions; (ii) farming – a combination of models from the livestock-only and mixed systems; and (iii) dominant – the mixed, humid and sub-humid region. These three maps are shown in Figure 13. As before, it is not implied that any of these represents the most appropriate model or resolution for estimating household expenditure.

The 'climate zones' and 'farming systems' regions cover the whole country (except for the 'urban' and 'other' farming systems), while the 'dominant' system map covers only the central area of Uganda. All three maps are similar to the SAE map of expenditure and to the OLS country-wide predictions. The 'climate zones' models capture more extreme ranges of expenditure than do the others, i.e. the northeast area of very low expenditure and the southwest area of very high expenditure. This is similar to the expenditure pattern seen in the SAE map (Figure 2).

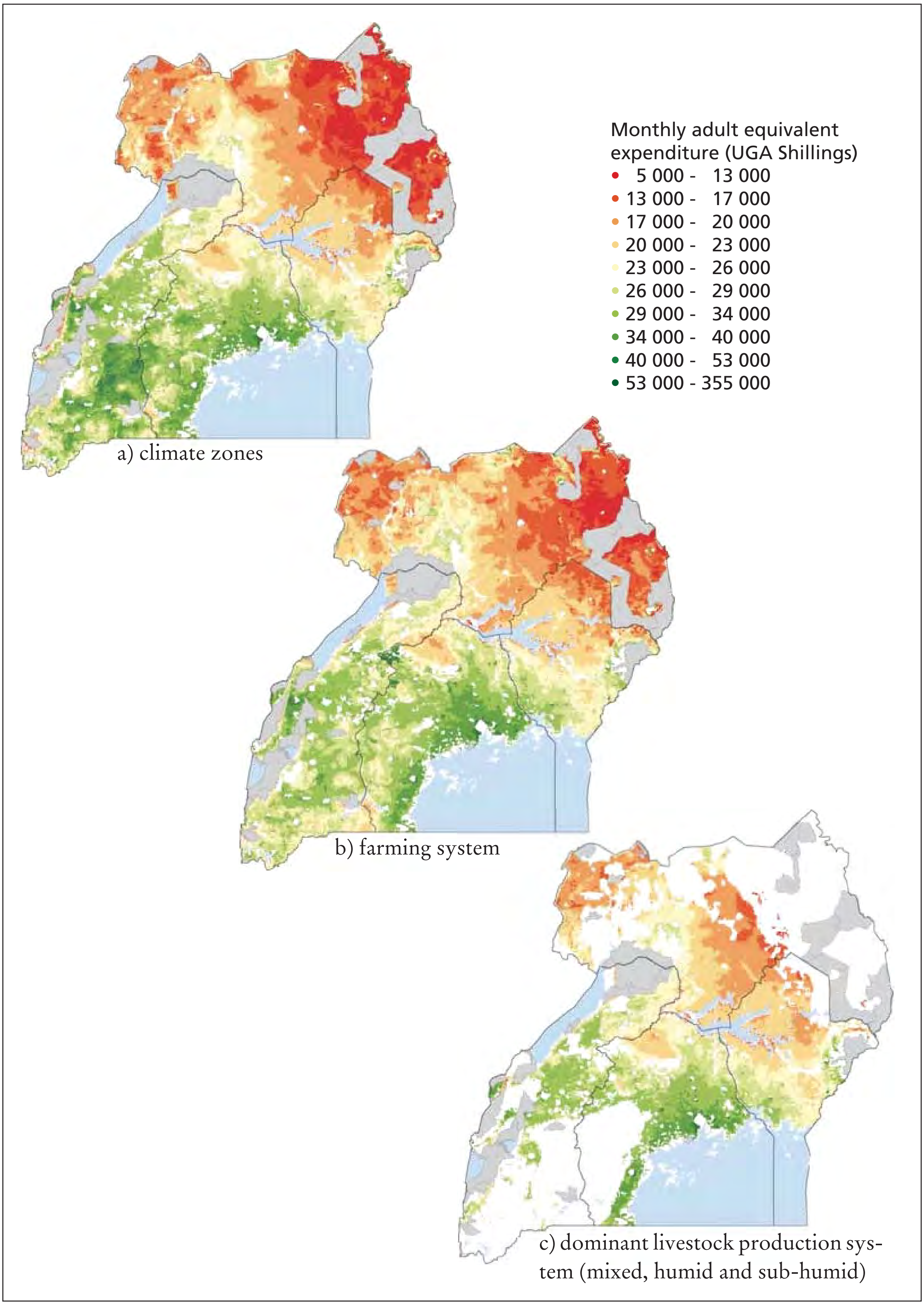
Table 8. Summaries for regional models at selected resolutions.

Model name and results			Significance§, sign (+/-) and relative importance (1-7)							
Model	Cell size	Points	R^2	ndvi	vpd	goat	cattle	slp	grump	Dist
C1 Arid system	0.01	296	0.310	*** + 1	- 3	- 5	*** + 4	+ 7	- 6	** - 2
	0.05	110	0.294	+ 3	- 2	* 5	+ 6	+ 7	+ 4	** - 1
	0.10	87	0.362	** + 2	- 3	** - 7	+ 5	- 6	+ 4	- 1
C2 Temperate system	0.01	1 404	0.174	*** + 2	*** - 1	*** - 5	- 7	*** - 6	*** + 4	*** - 3
	0.05	546	0.235	** + 2	*** - 1	** - 4	+ 7	*** - 5	*** + 3	*** - 6
	0.10	352	0.335	*** + 2	*** - 1	** - 4	+ 7	* - 6	*** + 3	- 5
C3 Humid & sub-humid system	0.01	291	0.183	+ 6	- 7	*** - 2	+ 4	- 5	+ 3	*** - 1
	0.05	99	0.322	+ 2	* + 4	- 3	- 7	** + 5	- 6	*** - 1
	0.10	68	0.340	+ 3	- 7	- 2	- 5	+ 6	- 4	*** - 1
F1 Livestock-only system	0.01	73	0.368	*** + 1	- 2	- 6	** + 3	+ 7	+ 4	- 5
	0.05	*25*	*0.390*	+ *4*	- *2*	- *5*	+ *1*	+ *3*	+ *6*	+ *7*
	0.10	*19*	*0.357*	+ *nd*	- *nd*	- *nd*	+ *nd*	- *nd*	+ *nd*	+ *nd*
F2 Mixed system	0.01	1 918	0.176	*** + 2	*** - 1	*** - 4	** + 7	** - 6	*** + 5	*** - 3
	0.05	730	0.216	*** + 2	*** - 1	*** - 3	* + 7	*** - 6	*** + 5	** - 4
	0.10	452	0.318	+ 1	- 2	- 4	+ 7	- 6	+ 5	- 3
Mixed, humid & sub humid system	0.01	1 357	0.168	*** + 2	*** - 1	*** - 4	+ 7	*** - 6	*** + 5	*** - 3
	0.05	531	0.231	** + 2	*** - 1	*** - 3	+ 7	*** - 5	*** + 4	+ 6
	0.10	343	0.323	*** + 2	*** - 1	*** - 4	+ 7	- 6	*** + 3	- 5

§ *** p<0.001 ; ** p<0.01 ; * p<0.05

Note: Most important variables in bold, table rows in italics signify possibly insufficient numbers of data points.

Figure 13. Predicted average rural monthly adult equivalent expenditure based on regional models at 0.01 degrees resolution (c. 1.1 km at the equator).



Note: Areas of no prediction are in white.

GWR RESULTS

The first stage of GWR calibrates the model using a cross validation approach to determine the best bandwidth or kernel size. The GWR model is then run at that bandwidth, producing the full range of outputs and tests of significance. The model is next applied to all rural pixels in Uganda to predict average rural monthly adult equivalent expenditure for each pixel, as before. Finally the coefficients and their significance levels are mapped and interpolated.

The results at 0.01 degree resolution are shown in Table 9. The first stage results indicated that the optimal kernel size should include 807 (38.7 percent) of the 2 088 data points available to develop a single regression model for each point.

Table 9 provides the quartiles of the distribution of each coefficient as it varied across the dataset, and also gives the OLS coefficient (named 'Global') for comparison. The results show that the GWR results do vary across the region with all coefficients (except the intercept) ranging from negative to positive.

The regression outputs, given as footnotes to Table 9, can be compared with the OLS results presented in Table 5 (though see notes of caution below, regarding the use of these internal statistics for comparing different models). The GWR model has a lower sigma, lower AICc and higher R^2 value than the country-wide OLS model. An ANOVA rejected the null hypothesis that the GWR model offered no improvement over the OLS model (F = 3.9959, df1 = 767.219, df2 = 2 046.027, p-value <2.2e-16).

The GWR model outputs make comparisons of the sign, significance and importance of variables between these and OLS models difficult, but the GWR procedure in R produces a series of statistics designed to compare the GWR model with the OLS model (Table 10): R^2 (the higher the better), AICc (the lower the better) and significance levels based on the p values from two different F-tests, F1 (Leung *et al.* 2000) and F2 (Fotheringham *et al.* 2002) that compare the GWR against the OLS model. The results at coarser resolutions are not presented because there are insufficient data points to describe the GWR kernels for them.

Table 9. Summary of GWR coefficient estimates at 0.01 degree resolution (c. 1.1 km at the equator), based on a kernel size of 807 data points (38.7 percent of the 2 088 data points available).

Variable	Min.	1st Qu.	Median	3rd Qu.	Max.	Global
(intercept)	4.85e+00	5.07e+00	5.24e+00	5.35e+00	5.61e+00	5.1555
ndvi	-4.04e-01	6.35e-02	1.63e-01	2.76e-01	6.11e-01	0.3085
vpd	-9.96e-02	-5.79e-02	-4.27e-02	-1.31e-02	4.85e-02	-0.0352
goat	-1.74e-03	-5.40e-04	-2.13e-04	3.75e-04	7.62e-04	-0.0007
cattle	-8.87e-04	-9.48e-05	3.34e-04	6.89e-04	1.07e-03	0.0004
slp	-2.69e-02	-1.64e-02	-7.17e-03	-4.67e-03	1.00e-02	-0.0058
grump	-9.46e-07	2.31e-05	4.59e-05	9.46e-05	1.78e-04	0.0001
dist	-3.05e-04	-1.55e-04	-7.83e-05	-5.02e-05	1.29e-05	-0.0001

Effective number of parameters: 55 97674; effective degrees of freedom: 2 032.023; sigma: 0.1031564; AICc: 3 524.839; residual sum of squares: 21 62324; GWR multiple R^2: 0.2555 (compared with the OLS multiple R^2 of 0.1852).

Both the R^2 and AICc scores suggest that GWR out-performs OLS across all resolutions, even when accounting for the added model complexity in GWR. However, statistics like these, estimated within a model, should really only be used in similar models, for example, the AICc is generally used to see if a particular predictor variable improves the fit of the model. Comparing R^2 and AICc values across models with different structures, different sets of variables and, most importantly, different numbers of data points is thus problematic. Consequently, in the following section, more robust comparisons are made between the different models.

GWR suffers from a lack of data points at resolutions above 0.05 degrees resolution. Figure 14 shows the predicted expenditure using the GWR model at 0.01 degrees resolution. Again, the resulting map is very similar to the results of the previous models, and to the SAE map.

Having run all models at all resolutions it was then possible to perform a direct comparison of the predictions across all models to identify the best performing model and resolution.

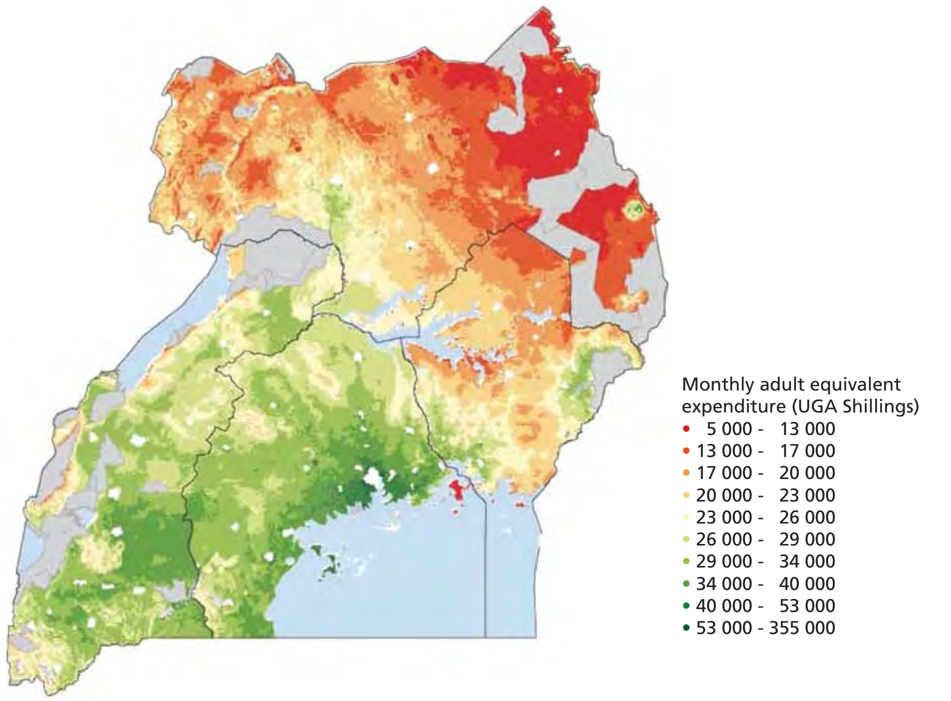
Table 10. GWR and OLS model comparison.

Model scale		Kernel size		R^2		AICc		Significant improvement?[1]	
Cell size	Points	Points	As a %	GWR	OLS	GWR	OLS	F1	F2
0.01	2 088	807	39%	0.256	0.185	-3 525	-3 411	***	***
0.02	1 279	419	33%	0.300	0.176	-363	-248	***	***
0.03	1 086	*201*	*19%*	*0.376*	0.206	*672*	762	***	***
0.05	813	387	48%	0.351	0.227	1 408	1 494	***	***
0.10	539	*178*	*33%*	*0.472*	0.292	*1 532*	1 599	***	***
0.15	399	*172*	*43%*	*0.460*	0.290	*1 962*	2 009	***	***

[1] *** $p<0.001$; ** $p<0.01$; * $p<0.05$

Note AICc scores should be compared at the same resolution, not across resolutions. GWR data in italics are based on few data points

Figure 14. Predicted average rural monthly adult equivalent expenditure based on the GWR model at 0.01 degrees resolution.



GOODNESS OF FIT METRICS FOR ALL REGRESSION MODELS AND THE SMALL AREAS ESTIMATES

Often R^2 or adjusted-R^2 values generated within regression models are used to compare models. However, when comparing regression models in which the dependent variable has been transformed in different ways, which used different sets of data points, and which include different combinations of independent variables then the model R^2 is not a reliable guide in comparing model quality. In such cases direct comparisons between the predicted values and the observations should be used, such at the R^2 estimate for the relationship between observed and model-predicted values, RMSE and other, related metrics.

Although the residual standard error (or Sigma) from a regression model is effectively the same as the RMSE, Sigmas cannot be compared directly across the models produced here because each model is based on a different transformation of the dependent variable. So, instead, after back transforming the predicted rural monthly adult equivalent expenditure for each of the *n* pixels containing rural households, the RMSE in Ugandan Shillings was estimated for each model at each resolution as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(predicted_i - actual_i)^2}{n}}$$

The mean absolute error (MAE) and mean absolute percentage error (MAPE) are two other measures of fit that are less sensitive to outliers than is the RMSE:

$$MAE = \frac{\sum_{i=1}^{n} |predicted_i - actual_i|}{n}$$

$$MAPE = \left(\frac{\sum_{i=1}^{n} |predicted_i - actual_i|}{actual_i} \right) \cdot \frac{1}{n}$$

Finally, for completeness, the R^2 value was computed from the plot of observed vs. expected expenditure for all data points, at all resolutions. However, this suffers from the same sensitivity to outliers as does the RMSE.

At each resolution the country-wide OLS, regional OLS and GWR models were bootstrapped. Each regression model was run 1 000 times with bootstrapped samples from the original dataset to obtain a distribution of the four metrics, which were then used to generate unbiased estimates and standard errors, shown in Table 11.

The same four metrics were estimated for the SAE expenditure maps at district, county and sub-county levels (Table 12). These could only be computed based on the administrative units that contained rural household points, just as the regression model used only those pixels that contained household points. The average administrative unit size (with standard errors) was estimated for each SAE, and two extreme outliers were removed from the sub-county level and two from the county level SAE results before computing the metrics[6].

Figure 15a shows the results for MAE, and Figure 15b, shows the same results in greater detail for the finer resolutions (up to 300 km^2, beyond which there were insufficient records to allow robust predictions to be made – see Table 11). Model performance is plotted against average pixel size in square kilometres, demonstrating the trade-off between model accuracy and the spatial resolution. The SAE results are also included on the graph, although it was not possible to compute standard errors for these (though standard errors around the average administrative unit area are given). In all cases the results for the regional OLS models lay between the country wide OLS and GWR results, though for clarity these have been omitted from Figure 15.

The results show that the GWR predictions were better than the regional OLS models, which, in turn, were better than the country-wide OLS. They also show that the country-wide OLS and GWR models have similar metric scores to the SAE models at cell sizes that were comparable to the district and country scales. However the OLS and in particular the GWR models had significantly better metric scores than the sub-county SAE models at comparable scales. For example, the sub-county RMSE was 16 614, comparable to the 0.02 degrees resolution GWR model, with an RMSE equal to 16 339; a 44-fold increase in spatial precision. For MAE and

[6] Sub counties 406206 and 205103 and their corresponding counties 4062 and 2051. There are no SAE for the corresponding districts 406 and 205.

MAPE the comparable GWR resolutions were 0.03 and 0.05 degrees; a 20- or 7-fold increase respectively, and for R^2 it was 0.01 degrees (a 178-fold increase)

Considering all the metrics in Table 11 and the shape of the curve in Figure 15b, a cell size of 0.05 degrees, covering approximately 31 km², or 5.5 × 5.5 km, results in a conservative trade-off between spatial precision and the predictive accuracy of the model. At this resolution (as with almost all others), GWR gives the best result followed by the regional OLS models for the dominant (mixed, humid and sub-humid) livestock production system, and finally the country-wide OLS model. Figure 16 shows the predicted average monthly rural household expenditure for the GWR model at 0.05 degrees resolution. These estimates have lower or comparable errors to the finest SAE map and are over seven times as detailed as the SAE rural monthly adult equivalent expenditure estimates at sub-county level. The summary results for the 0.05 degree GWR model are also shown.

Table 11. Goodness of fit metrics for GWR and OLS models at each resolution.

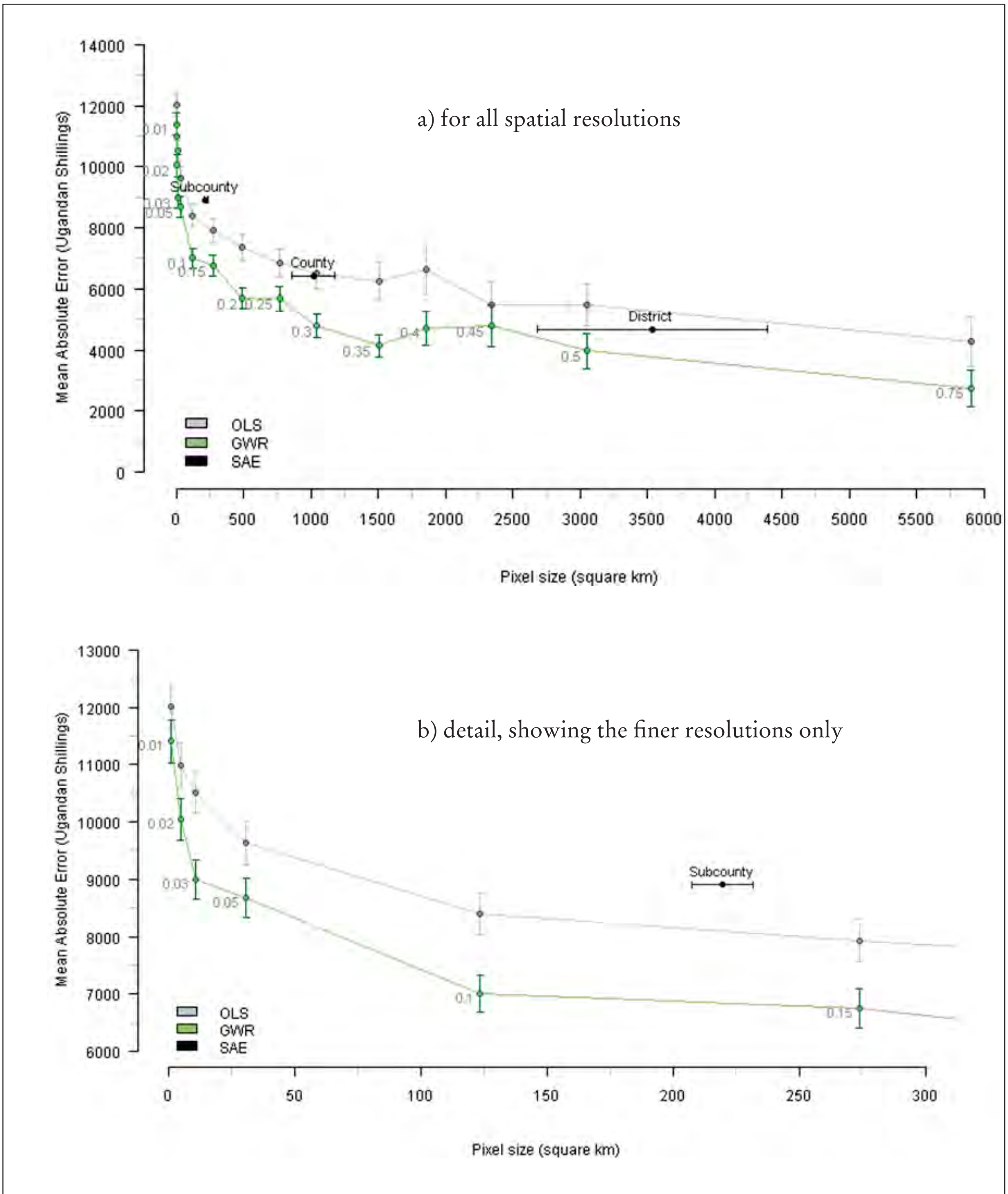
			RMSE (UGA Shillings)		MAE (UGA Shillings)		MAPE (%)		R^2 (for Obs vs Exp)	
Cell size (degrees)	Records	Sq km	GWR	OLS	GWR	OLS	GWR	OLS	GWR	OLS
0.01	2 088	1.2	20 462±1 563	21 034±1 531	11 408±371	12 024±382	37.4±0.7	40.2±0.8	0.17±0.02	0.11±0.02
0.02	1 279	4.9	16 339±953	17 563±968	10 044±368	10 991±387	33.5±0.9	37.7±1.0	0.25±0.03	0.13±0.02
0.03	1 086	11.1	14 053±784	16 091±824	8 996±333	10 518±355	30.4±0.9	36.5±1.0	0.37±0.03	0.17±0.03
0.05	813	30.9	12 893±713	14 173±719	8 680±348	9 637±374	30.6±1.0	34.1±1.1	0.34±0.04	0.22±0.04
0.10	539	124	9 866±602	11 772±644	7 001±316	8 394±366	23.8±0.9	29.3±1.2	0.51±0.04	0.30±0.04
0.15	399	274	9 170±455	10 854±520	6 746±340	7 933±378	24.1±1.4	29.0±1.5	0.51±0.04	0.32±0.05
0.20	*280*	*493*	*7 690±433*	*9 840±572*	*5 700±354*	*7 347±426*	*20.5±1.3*	*27.3±1.7*	*0.64±0.05*	*0.40±0.07*
0.25	*206*	*770*	*7 660±498*	*9 236±576*	*5 680±410*	*6 846±477*	*21.9±2.0*	*27.0±2.6*	*0.61±0.06*	*0.43±0.08*
0.30	*167*	*1 047*	*6 492±519*	*8 469±577*	*4 799±384*	*6 490±496*	*17.1±1.5*	*23.7±2.1*	*0.69±0.04*	*0.46±0.06*
0.35	*120*	*1 504*	*5 484±424*	*8 646±976*	*4 132±368*	*6 266±630*	*14.7±1.6*	*22.9±2.6*	*0.83±0.04*	*0.58±0.08*
0.40	*103*	*1 854*	*6 500±747*	*9 108±1 166*	*4 727±555*	*6 622±782*	*17.6±2.4*	*25.7±3.3*	*0.72±0.05*	*0.46±0.08*
0.45	*82*	*2 342*	*6 731±1 080*	*7 625±1 243*	*4 784±678*	*5 500±734*	*16.4±2.5*	*18.6±2.7*	*0.76±0.08*	*0.69±0.09*
0.50	*75*	*3 051*	*5 561±884*	*7 478±976*	*3 960±579*	*5 499±687*	*16.2±2.6*	*23.1±3.5*	*0.78±0.08*	*0.61±0.10*
0.75	*36*	*5 903*	*3 743±730*	*5 386±903*	*2 753±599*	*4 283±830*	*12.6±3.6*	*19.3±4.8*	*0.87±0.05*	*0.72±0.08*

Note: Rows in italics are models with few data points.

Table 12. Goodness of fit metrics for the Small Area Estimates at each scale.

SAE unit	Records	Sq km	RMSE	MAE	MAPE (%)	R^2
Sub County	528	220±12	16 614	8 910	29.9	0.14
County	144	1 018±159	9 109	6 432	20.7	0.49
District	53	3 537±859	6 153	4 669	17.4	0.68

Figure 15. Mean Absolute Error, with bootstrapped standard errors over 1 000 replications, for country-wide GWR and OLS regression models at all resolutions, and for SAE.



Note: GWR points are labelled with the cell size in degrees. The horizontal error bars on the SAE values show the standard errors of the mean area of the administrative units.

Figure 16. Predicted average rural monthly adult equivalent expenditure based on the best performing method, a Geographically Weighted Regression (bandwidth = 387 neighbours) model at 0.05 degrees resolution (c. 5.5 km at the equator).

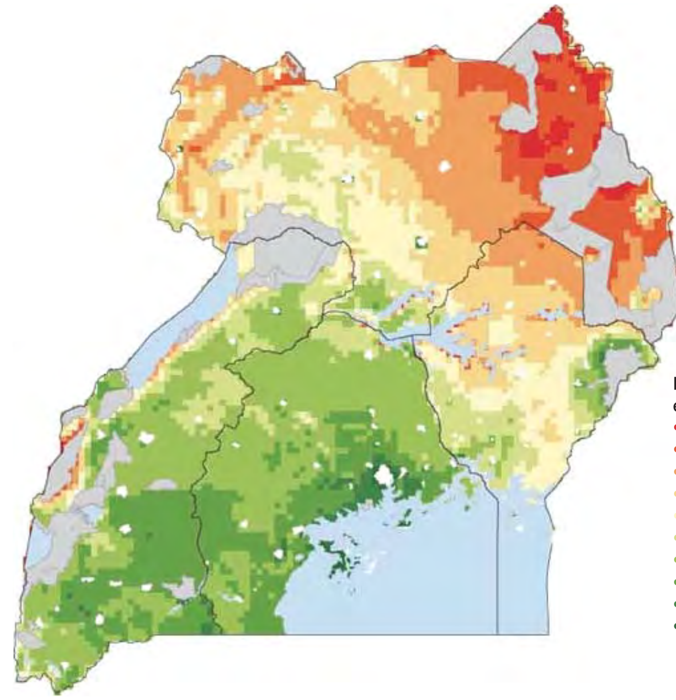


Table 13. Summary of GWR coefficient estimates at 0.05 degree resolution (c. 5.5 km at the equator), based on a kernel size of 387 data points (47.6 percent of the 813 data points available).

Variable	Min.	1st Qu.	Median	3rd Qu.	Max.	Global
(intercept)	1.12e+01	1.21e+01	1.27e+01	1.31e+01	1.37e+01	2.8100
ndvi	-9.46e-01	2.37e-01	5.34e-01	9.74e-01	1.71e+00	0.7700
vpd	-5.29e-01	-3.51e-01	-2.23e-01	-9.03e-02	2.63e-01	-0.2994
goat	-1.15e-02	-6.16e-03	-1.23e-03	3.53e-03	8.31e-03	-0.0062
cattle	-6.14e-03	-1.21e-03	2.77e-03	4.21e-03	7.07e-03	0.0023
slp	-6.61e-01	-3.17e-01	-1.55e-01	-6.22e-02	1.77e-01	-0.1149
grump	1.21e-04	3.75e-04	5.58e-04	6.81e-04	1.90e-03	0.0006
dist	-8.04e-04	-2.28e-04	2.97e-05	1.59e-04	6.57e-04	-0.0001

Effective number of parameters: 43.80606; effective degrees of freedom: 769.194; sigma: 0.5493839; AICc: 1 367.77; residual sum of squares: 245.3818; GWR multiple R^2: 0.3514 (compared with the OLS multiple R^2 of 0.2271).

SPATIAL VARIATION IN THE GWR COEFFICIENTS

This section explores whether significant spatial variation in the GWR coefficients was present. Such variation would imply that the dependent variables relate to rural monthly adult equivalent expenditure in different ways in different areas of Uganda. In extreme cases, strong variation may indicate model misspecification (i.e. the need to use different dependent variables in a particular location). Although such variation can be investigate at a range of spatial resolutions and bandwidths, here, the analysis is presented only for the 'best' model; at 0.05 degrees resolution with a bandwidth of 387 neighbours.

Leung *et al.* (2000) developed a formal F test for GWR to determine if the variation in the GWR coefficients is significant. The results (Table 14), given below, show that NDVI, with p = 6.6 percent, fell short of significant at the usually accepted 5 percent level, but that all the other coefficients were significant at the 0.1 percent level or better. In other words there is significant spatial variation in most of the GWR coefficients.

Table 14. Test for spatial variation in the GWR coefficients based on the method.

Variable	F statistic	Numerator d.f.	Denominator d.f	Pr(>)	Significance[1]
(intercept)	2.5323	86.5142	780.28	2.182e-11	***
ndvi	1.3833	37.1461	780.28	0.06611	(*)
vpd	4.8308	279.4226	780.28	<2.2e-16	***
goat	4.7707	292.2597	780.28	<2.2e-16	***
cattle	3.6694	213.6288	780.28	<2.2e-16	***
slp	2.7149	81.4713	780.28	1.548e-12	***
grump	3.0879	24.9101	780.28	7.875e-07	***
dist	1.9658	145.0732	780.28	5.383e-09	***

[1]*** p<0.001 ; ** p<0.01 ; * p<0.05; (*) p<0.1.

Based on this the coefficients were mapped, along with their significance levels (based on a t-test). The maps in Figure 17 are arranged in pairs with the coefficient maps presented alongside the significance maps, for each parameter in turn, showing the following.

- The spatial variation in each coefficient using a red-blue (low-high) bipolar colour scheme based on standard deviations of the coefficient values.
- The zero value (where it exists) of the GWR coefficient as a green 'contour' line, to demarcate where the coefficient switches from a positive to a negative effect. Negative areas are generally in red shades, but population density is the exception where there are no negative values.
- The country-wide OLS parameter value as a black 'contour' line.
- The regions where the coefficients are significant are shaded in green; the darker the shade, the higher the level of significance.

Unlike the previous maps, protected areas, urban areas and lake overlays are not shown, as it would further complicate the maps, without aiding interpretation. The regional boundaries are given as a locational aid.

Figure 17. Mapped coefficient values (left-hand side) and their significance levels (right hand side) for the 0.05 degrees resolution GWR model.

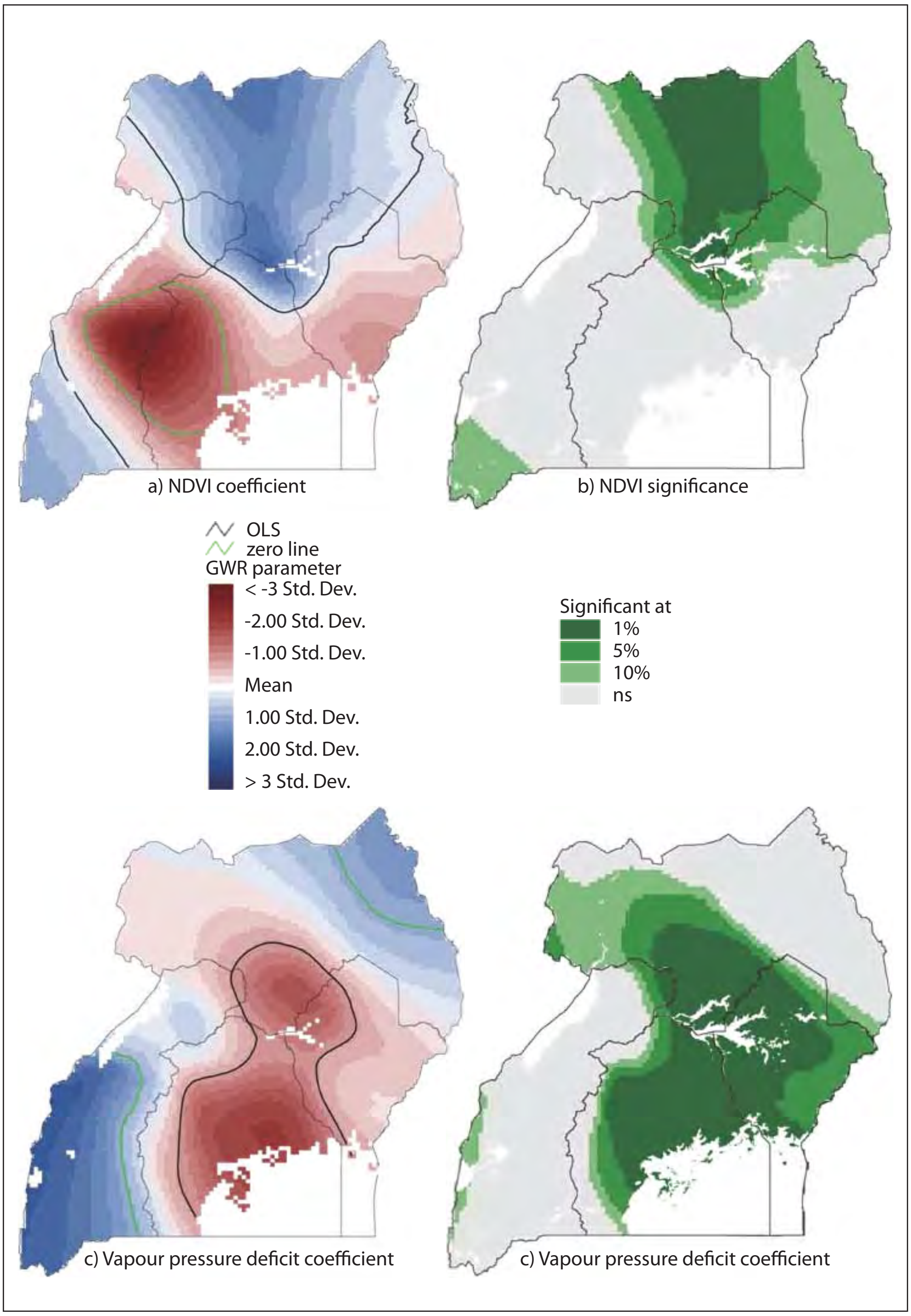


Figure 17. Continued.

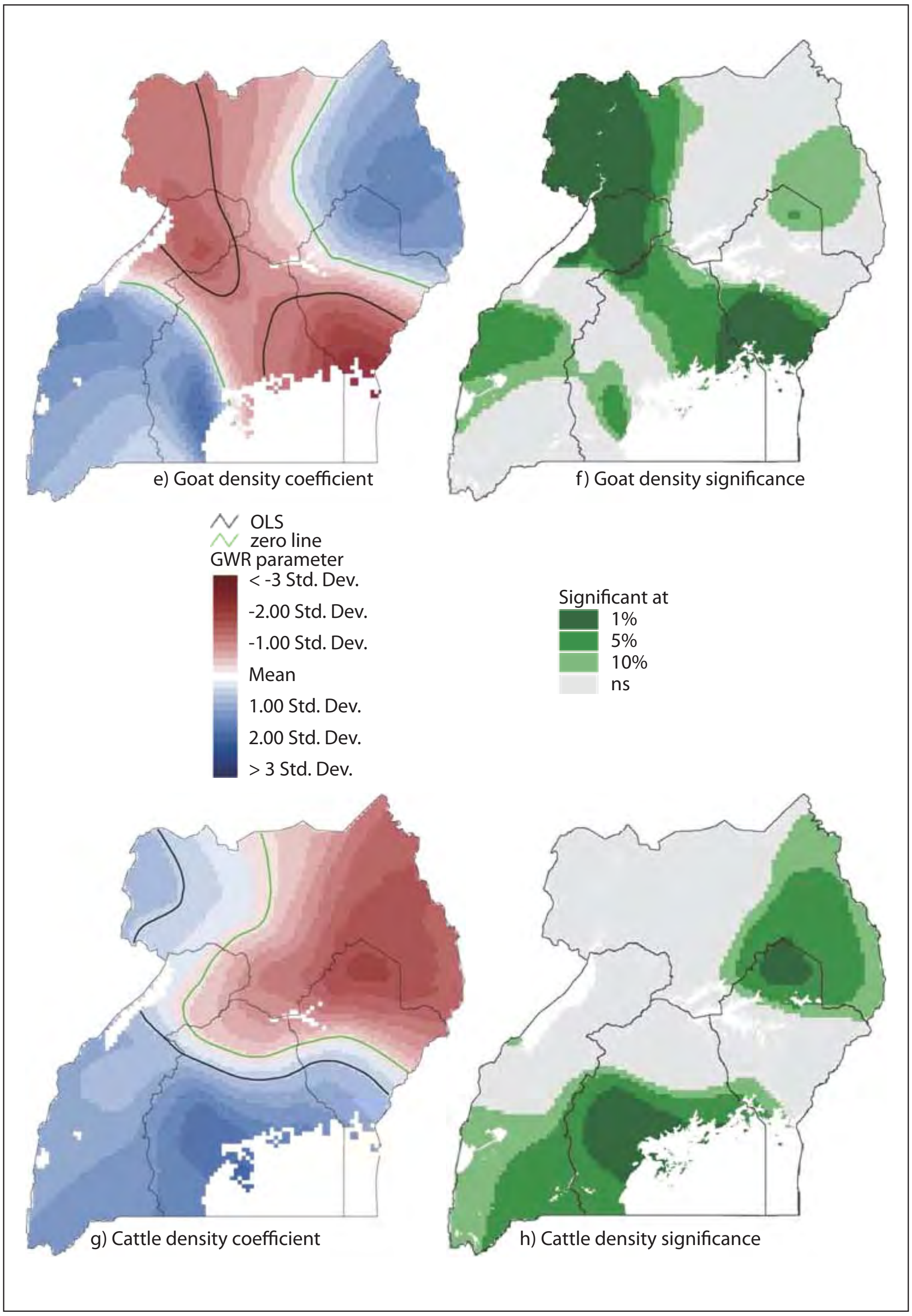


Figure 17. Continued.

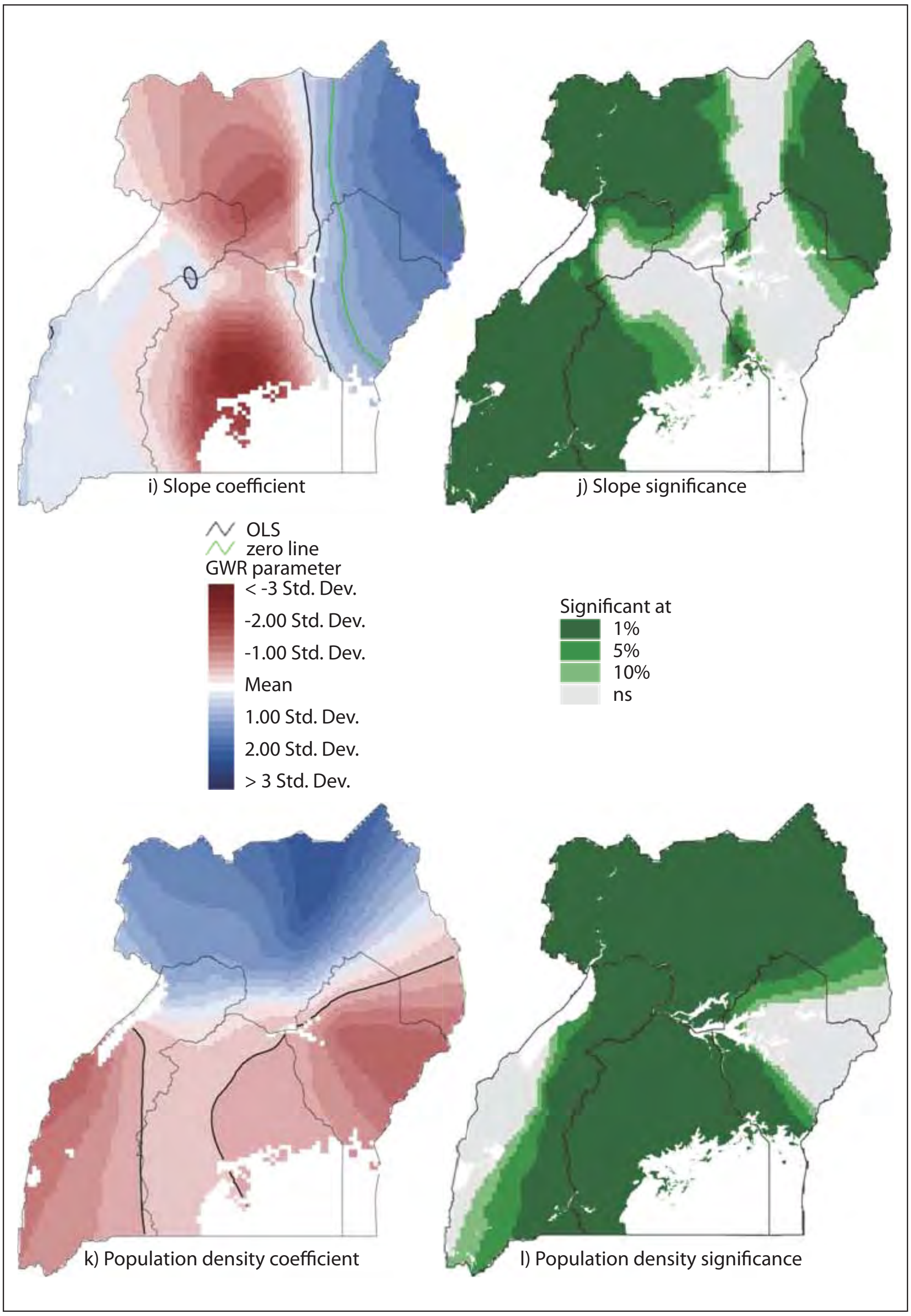
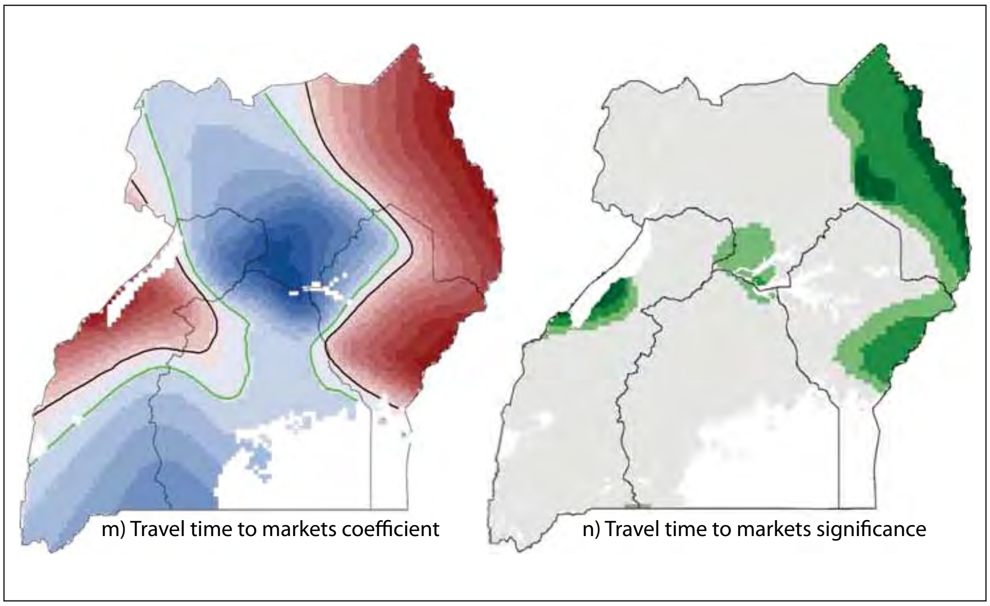


Figure 17. Continued.



INTERPRETING THE GWR COEFFICIENTS

In this section, possible explanations are offered for the observed patterns in each GWR parameter map, and its attendant significance estimate, in an attempt to interpret what the resulting patterns may say about how rural poverty is related to environmental conditions in different areas. Whilst we talk here about positive or negative regression coefficients we do not mean to imply the causation usually associated with regression results. The results still only confirm correlation rather than causation.

NDVI

The expected influence of NDVI would be positive, with greater vegetation vigour corresponding to higher levels of expenditure. The results (Figures 17a and 17b) showed strong positive correlations in all areas other than a patch in the centre/southwest of the country (within the green contour of Figure 17a).

Higher NDVI values broadly indicate richer vegetation growth, longer growing season(s) and higher rainfall. In the drier areas in the north, northeast and extreme southwest the coefficient of NDVI values on expenditure is positive, as one would expect, and it is in these areas that the GWR parameter is significant (according to the t-test). In the much greener areas of the central part of Uganda, the NDVI coefficient is not significant so the reversal of its influence in these areas is not of concern.

These patterns suggest that in this model there is a saturation level in terms of vegetation vigour, beyond which there is little or no benefit. Large areas of Uganda are well-served in terms of length of growing period (which is highly correlated with the annual integrated NDVI). It is only in the relatively dry areas that NDVI is likely to be limiting for agricultural production and thus to livelihood options and welfare.

Vapour Pressure Deficit

The expected influence of VPD was negative; the greater the deficit the lower expenditure would be. The results (Figures 17c and 17d) showed the relationship to be negative, except for in the extreme northeast and southwest regions of the country (beyond the green contours of Figure 17d).

VPD is a measure of the drying power of air and, where it has a significant influence, its coefficient against household expenditure is negative (i.e. the lower the VPD the higher the expenditure): in the north/central area of the country, above Lake Kyoga and the northwest shores of Lake Victoria. In the more arid northeast and the very humid southwest regions, where the sign of the coefficient is the reversed, VPD is not significant.

In the areas where the VPD coefficients are significant, the average VPD values are relatively low so, one might expect, not be limiting to welfare. It seems that in the drier areas of the northeast, where VPD is much higher and possibly more limiting to agricultural development and livelihood options, other variables are coming into play. NDVI in particular, is significant in these areas so presumably better accounts for the aridity. Also, the very different agricultural systems in central and northeast Uganda may be differentially affected by VPD (and NDVI)

Goat density

The expected influence of goat densities is ambiguous; in areas where goats are kept a positive effect might be expected (higher goat densities corresponding to higher expenditure), though in general goats are only kept in the more arid and isolated pastoral areas so are possibly indicative of lower average levels of welfare. The results (Figures 17e and 17f) showed a distinct northeast to southwest trend; with negative sign in the central region, and positive sign at either end of that trend (beyond the green contours of Figure 17f).

There are three distinct areas of significance in the goat coefficient: (i) in the arid, northeast pastoral areas, where goat densities are at their highest, and their influence on welfare is strongly positive; (ii) in a central band (northwest to southeast), again of relatively high goat density, where their influence is negative; and (iii) an area in the southern part of the country, where goat densities are lower, and their influence is again positive. The negative effects in the flatter more humid central regions may well reflect the variation in growing conditions within that zone; goats are likely to be raised in the drier areas less suited to cropping which would give rise to them being associated with lower welfare levels in these otherwise productive and relatively affluent regions. Goat density is positively related to expenditure in the more arid pastoral areas in the northeast: in these generally poorer areas goats are of great importance and are likely to reflect relative wealth. The positive influence in the temperate southwest is difficult to explain.

Cattle density

The expected influence of cattle density was positive, with higher densities giving rise to, or reflecting, greater wealth (expenditure). The results (Figures 17g and 17h) again showed a distinct northeast to southwest trend; with negative influence to the northeast of the green contour of Figure 17h), and positive influence southwest of it.

There are two areas in which the influence of cattle density was significant: (i) in the pastoral northeast, where the influence was negative, and (ii) in the southern part of the country, where the influence was positive. This latter result is easily explained: these are important cattle areas and dairy production, in particular, is prevalent. Why cattle densities should have a negative influence on expenditure in the northeast is difficult to fathom: one would expect a pattern similar to that of goats, whereby larger numbers reflect greater wealth among the pastoralists. It could be argued that too much should not be read into the coefficient map, however, since the parameter does not contribute strongly to the OLS model (c.f. Table 7 and Figure 14), and there are few areas where the GWR parameter is significant at the 5 percent level. Nonetheless, the pattern is intriguing and deserves further investigation.

Slope

The expected influence of slope was negative; steeper slopes corresponding to lower expenditure. The results (Figures 17i and 17j) indicated a strong east-west pattern; to the east of the zero contour (shown in green in Figure 17j there is a positive influence, whilst west of this contour the influence is negative.

The influence of slope is significant in three areas of Uganda: (i) the arid northeast, where its influence is positive; (ii) the northwest where its influence is negative, as expected; and (iii) the southwest of the country, where its influence again is negative, increasingly so from west to east and most strongly so close to the shores of Lake Victoria. Where it is significant, therefore, its influence is negative in the mixed farming areas, which is to be expected, since rough terrain hinders cultivation. Slope is less important in areas dominated by livestock, such as the northeast; the significant positive slope in this region therefore does not contradict the conclusions drawn from the other regions.

Population density

The expected influence of population density was positive; the greater the density of people the higher expenditure would be. The results (Figures 17k and 17l) revealed a strong north-south pattern; more positive in the north and less positive in the south of the country, but with no regions in which the influence of population density on expenditure was negative (there is no green, zero contour in Figure 17k).

The influence of population density is significant over most of the country, the exceptions being the southwest border, and a curiously-shaped wedge, fanning out to the east of Lake Kyoga. Both of these areas are where the population density coefficients are at their lowest values and, incidentally, they coincide with areas of steeply sloping land: Mount Elgon in the east and the Ruwenzori Mountains to the west. The population coefficient is more positive in the north than in the south, suggesting that the influence on expenditure of population density is less in areas of high rather than low population density. This may point to a saturation effect, in that there are diminishing returns to being near or in a high density area above a certain density threshold.

Travel time to markets

The expected influence of travel time to markets is negative; in general poor areas tend to be remote and higher expenditure would be expected in areas with good market access (with quick access to markets). The results (Figures 17m and 17n) showed bimodal, east-west trend; with negative sign in the more remote west and in the eastern parts of the country, beyond the green contours of Figure 20m, and positive sign within those contours, in the central and southwest parts of the country.

Travel time to markets was, perhaps surprisingly, one of the least influential parameters in the OLS regression model (c.f. Table 7 and Figure 11) and there are few areas where the GWR parameter is significant (Figure 17n), which makes interpretation difficult. There are three areas of significance: the eastern border of the country, where market access is poor and the influence of travel time is negative, as expected; (ii) a small area to the west, on the shores of Lake Albert, where market access is again generally poor and the influence of travel time is again negative, as expected; and (iii) a small area to the west of Lake Kyoga (only significant at the 10 percent level), where there is a surprising positive influence of travel time on expenditure. The patterns in the east and west suggest that increased access to markets for the more isolated regions would be beneficial. Interestingly though, this variable does not have a significant effect in all areas that are far from markets: the northwest, for example.

Conclusions

The validity of an environmental approach to poverty mapping was clearly demonstrated by Rogers *et al.* (2006) and Robinson *et al.* (2007), who used discriminant analysis of Fourier-processed multitemporal satellite data, combined with other relevant environmental variables, to model and predict household expenditure in Uganda. They presented an approach to mapping poverty that took us beyond description, where the more traditional small area estimates reach their limits, and towards explaining the distribution of poverty, and possibly predicting changes in poverty that may result from changing, through careful intervention, the conditions observed to be associated with it.

Here, that analysis has been taken a step further, using regression techniques that are readily accessible from routines within the R environment for statistical evaluation. This makes the analysis performed here readily reproducible. Three levels of spatial disaggregation have been investigated: global and regional analyses using ordinary least squares regression and a geographically weighted regression. As would be expected, dividing the area into zones, livestock production systems in this case, prior to regression improves the predictive power considerably. Because a zonation highly relevant to the role of livestock in agriculture and poverty alleviation was used, the relationships between poverty and the environment have been separately elucidated in these different livestock production systems; indicating which factors are most closely related to poverty in the different systems.

Only 7 of the predictor variables that were used in the original analysis (Rogers *et al.* 2006) were used, chosen largely on the grounds of avoiding variables that were highly correlated with others. This was done with the intention of getting a better understanding of the nature and relative importance of key variables at different levels of detail and using different zonations. For example, VPD and population density were consistently the two most influential factors in the OLS model at different resolutions and yet when the livestock production system zones were taken into account NDVI became more important in livestock only and in arid and semi-arid production systems. Importance is more difficult to assess in the GWR model but the maps of the GWR coefficients and their significance levels suggest that there is considerable spatial variation in the influence of factors like slope, population density and VPD.

The present analysis was restricted to rural households, and used per-adult equivalent expenditure (rather than total household expenditure) as the dependent measure of poverty. This has enabled a direct comparison of the environmentally-based results against those from the more traditional SAE approach (Emwanu *et al.* 2007), which have become available since the original (Rogers *et al.* 2006) analysis was conducted. This comparison has shown that an environmental approach to poverty mapping in Uganda consistently out-performs SAE approaches at equivalent spatial resolutions.

The 'best' performing model was the GWR at 0.05 degrees resolution (c. 5.5 km at the equator). There is a case for using the 0.03 degree resolution data to give a 20 fold increase in spatial resolution over the finest SAE map. However, caution has been exercised in the presentation of results here, and the door left open for better and finer resolution maps, which could take advantage of other environmental variables and more appropriate regional models, as indicated by the spatial patterns in the GWR parameter maps.

With respect to the SAE methodology, the disadvantage of the environmental approach is that the predictions are not made at the level of the household, so it is not possible to compute aggregate measures such as head counts and Gini indices. Nevertheless, the approaches demonstrated here, and in Rogers *et al.* (2006), have a role to play in understanding the nature of the relationships between poverty and socio-economic and environmental factors. It is not suggested that these models can identify causal links between poverty and the environment but they do form part of an accumulation of evidence that strongly suggests that spatial patterns of poverty, and possibly spatial poverty-traps, can be partially explained by environmental factors. This knowledge should lead to spatially-targeted policy support for poverty alleviation.

The GWR results show significant spatial variation and suggest that other zoning systems should be considered when designing statistical approaches to modelling and mapping poverty. One option for visualising these zones has been briefly demonstrated, to reveal regions that have similar coefficients in the model; where the relationship between poverty and the environment are consistent.

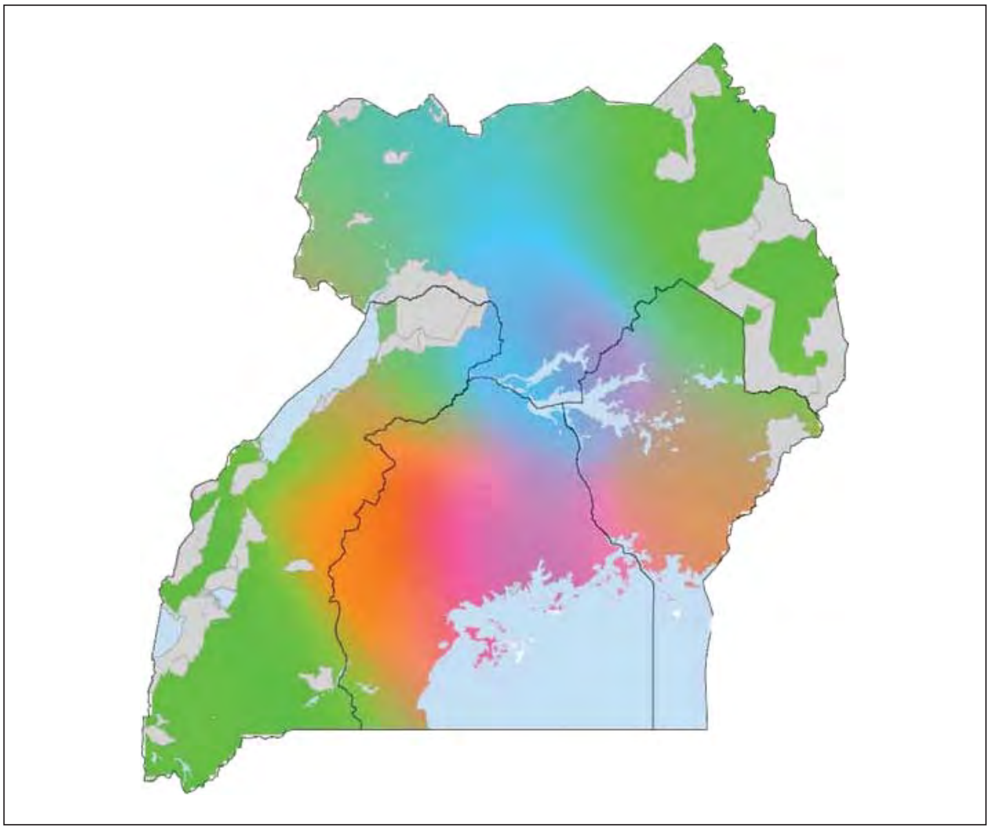
Figure 18 was derived from the GWR coefficient maps by first reducing the dimensionality of the data from eight (seven coefficients plus the intercept) to two, using a non-linear dimension-reduction technique called Sammon mapping (Sammon 1969), although Principal Components Analysis would also have served as a linear dimensional reduction technique. The purpose of dimension reduction is to reduce the N (8) variables to n (2) independent and orthogonal components that represent the maximum amount of information in the original N variables demonstrable in only two dimensions.

Each pixel was assigned a colour based on these two components using the CIELAB colour space which was three axes, L - lightness, a – hue and, b – chroma (CIE 1976). CIELAB is a unique colour space in that the distance between colours is perceptually uniform so there is a direct correspondence between distance or dissimilarly between data points and their assigned colours. Other colour spaces such as RGB do not have this property, which makes interpretation of RGB composite images challenging. In Figure 21 the two components have been assigned to the a and b CIELAB elements, and L held constant.

Figure 18 represents the multivariate spatial structure of the coefficients. The strong spatial patterns in the colour coding suggest that it is possible to use GWR to identify suitable zones of analysis based on the spatial relationship between poverty and its possible determinants. This is a very different approach to the more usual clustering and typologies that can be derived from the input variables, because the typologies here have been derived based on the *influences* that these variables may have on poverty, not based on the values of the variables themselves.

The colour coding serves merely to distinguish among different clusters of relationships between expenditure and environmental variables – the map cannot be interpreted beyond that; green is not 'better', or less poor, than red, for example. This simple linear combination must, however, be treated with caution since, ideally, each parameter should be weighted depending on its local significance. The cluster map presented in Figure 18 does, however, reveal intriguing spatial patterns that should be explored further.

Figure 18. Colour coded composite map of the GWR parameters at 0.05 degrees resolution



All of the models presented here are linear. The original discriminant analysis approach to this same dataset (Rogers *et al.* 2006; Robinson *et al.* 2007) employed an essentially non-linear technique, although it suffered from the constraint of discriminant analysis that the continuous expenditure data had to be binned into a number of expenditure categories before analysis. Descriptions and predictions were made based on the co-variance matrices of the key predictors best able to separate the different categories. This sort of discriminant analysis quite flexibly describes many different sorts of non linearities, as would alternative flexible approaches such as generalised additive models (GAMs). Alternatively the present linear models could incorporate certain sorts of non-linearities through the use of transforms (e.g. squares, square roots) of the descriptor variables, although these are fairly restrictive and must be specified in advance of modelling, rather than during it.

Many new questions are posed by this analysis. Is there scope for combining these kinds of environmental-geographical models with the census-survey data approach as used in the development of Small Area Estimate poverty maps? Can GWR be used to suggest zonations for different SAE models? Should environmental variables be used more commonly directly within the SAE methodology? These are all issues that should be explored further (i) to extract as much useful information as possible out of detailed spatial datasets, (ii) to develop more refined poverty estimates and, most importantly, (iii) to better understand the spatial patterns of rural poverty, and how these patterns relate to the environment.

Rogers *et al.* (2006) concluded '*what we have been able to show here is the step beyond exploiting correlations within internally correlated socio-economic data sets (the traditional small area mapping approach) to a situation where we have been able to show that external, independent data appear to have at least as much descriptive power for poverty mapping. The precise interpretation of the correlations obtained here will require more research effort but at least we have shown that this effort is both justified and appropriate.*' The work presented in this paper reinforces the justification of the environmental approach and takes some steps further towards explaining the pattern of poverty in Uganda.

References

Alderman, H., Babita, M., Lanjouw, J., Lanjouw, P., Makhatha, N., Mohamed, A., Ozler, B. & Qaba, O. 2000. *Is census income an adequate measure of household welfare? Combining census and survey data to construct a poverty map of South Africa*. Washington DC: Statistics South Africa and the World Bank.

Bollen, K.A. & Jackman, R. 1990. Regression diagnostics: An expository treatment of outliers and influential cases. In: J. Fox & J. Scott Long (eds.) *Modern methods of data analysis*, pp. 257-291. Newbury Park: Sage.

Box, G.E.P. & Cox, D.R. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26**:** 211-252.

Chomitz, K.M. & Gray D.A. 1996. Roads, land use, and deforestation: A spatial model applied to Belize. *World Bank Economic Review* 10: 487-512.

CIE. 1976. *Colorimetry*. CIE Publication 15.2, Commission International de L'Eclairage, Vienna.

CIESIN. 2004. *Global Rural-Urban Mapping Project (GRUMP), Alpha version: Population density grids*. Socioeconomic Data and Applications Center (SEDAC), Columbia University.

Davis, B. 2003. *Choosing a method for poverty mapping. Poverty and Food Insecurity Mapping Project report*. Rome: Food and Agriculture Organization (FAO).

Draper, N. & Smith, H. 1988. *Applied Regression Analysis*. New York: John Wiley and Sons.

Elbers, C. & Lanjouw, P. 2000. Intersectoral transfer, growth, and inequality in rural Ecuador. *World Development* 29: 481-496.

Emwanu, T., Okwi, P.O., Hoogeveen, J.C., Kristjanson, P. & Henninger, N. 2007. *Nature, Distribution and Evolution of Poverty and Inequality in Uganda*. Kampala: Uganda Bureau of Statistics (UBOS).

Emwanu, T., Okwi, P.O., Hoogeveen, J.G. & Kristjanson, P. 2003. *Where are the poor? Mapping patterns of well-being in Uganda, 1992-1999*. Kampala: Uganda Bureau of Statistics. 90 pp.

Epprecht, M. & Heinimann, A. 2004. *Socioeconomic atlas of Vietnam: A depiction of the 1999 Population and Housing Census*. Hanoi and Berne: NCCR North-South.

Fan, S., Zhang, X., & Rao, N. 2004. Public expenditure, growth, and poverty reduction in rural Uganda. Washington, DC: International Food Policy Research Institute (IFPR), Development Strategy and Governance Division. Discussion paper 4.

FAO. 1996. *World livestock production systems: current status, issues and trends*, by C. Seré and H. Steinfeld. Rome: FAO Animal Production and Health Paper 127. Rome: Food and Agriculture Organization (FAO) of the United Nations, Animal Production and Health Division.

FAO. 2007. *Gridded Livestock of the World, 2007*, by G.R.W. Wint and T.P. Robinson. Rome: Food and Agriculture Organization (FAO) of the United Nations, Animal Production and Health Division.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. London: John Wiley and Sons.

Garnick, D.H. 1990. Accounting for regional differences in per capita personal income growth: an update and extension. *Survey of Current Business* 70: 29-40.

Ghosh, M. & Rao, I.N.K. 1994. Small area estimation: An appraisal. *Statistical Science* 9: 55-93.

Griffith, D. and Amrhein, C.G. 1997. *Multivariate statistical analysis for geographers*. Upper Saddle River, NJ: Prentice Hall.

Grömping, U. 2007. Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician* 61: 139-147.

Hay, S.I., Randolph, S.E. & Rogers, D.J. 2000. *Remote sensing and geographic information systems in epidemiology*. London: Academic Press.

Hentschel, J., Lanjouw, I.O., Lanjouw, O. & Poggi, I. 2000. Combining census and survey data to trace the spatial dimensions of poverty: A case study for Ecuador. *The World Bank Economic Review* 14: 147-165.

Kruska, R.L., Reid, R.S., Thornton, P.K., Henninger, N. & Kristjanson, P.M. 2003. Mapping a livestock-oriented agricultural production system classification for the developing regions of the world. *Agricultural Systems* 77: 39-63.

Leung, Y., Mei, C.L. & Zhang, W.X. 2000. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A* 32**:** 9-32.

Lindeman, R.H., Merenda, P.F. & Gold, R.Z. 1980. *Introduction to bivariate and multivariate analysis*. Glenview IL: Scott, Foresman.

Minot, N., Baulch, B. & Epprecht, M. 2003. *Poverty and inequality in Vietnam: Spatial patterns and geographic determinants*. Hanoi: International Food Policy Research Institute (IFPRI) and Institute for Development Studies (IDS).

Ndeng'e, G., Opiyo, C., Mistiaen J. & Kristjanson, P. 2003. *Geographic dimensions of wellbeing in Kenya. Where are the poor?* Volume 1. Nairobi: The Regal Press Kenya Ltd. 164 pp.

Pfeiffer, D., Robinson, T., Stevenson, M., Stevens, K., Rogers, D. & Clements, A. 2008. *Spatial analysis in epidemiology*. Oxford: Oxford University Press. 142 pp.

Pozzi, F. & Robinson, T. 2008. Poverty and welfare measures in the Horn of Africa. IGAD Livestock Policy Initiative Working Paper number 8. Rome: Food and Agriculture Organization of the United Nations (FAO).

Pozzi, F., Robinson, T. & Nelson, A. 2009. Accessibility mapping and rural poverty in the Horn of Africa. IGAD Livestock Policy Initiative Working Paper number 40. Rome: Food and Agriculture Organization of the United Nations (FAO).

R-Development-Core-Team. 2009. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reuter, H.I., Nelson, A. & Jarvis, A. 2007. An evaluation of void filling interpolation methods for SRTM data. *International Journal of Geographic Information Science* 21**:** 983-1 008.

Robinson, T.P., Thornton P.K., Franceschini, G., Kruska, R.L., Chiozza, F., Notenbaert, A., Cecchi, G., Herrero, M., Epprecht, M., Fritz, S., You, L., Conchedda, G. & See, L. 2011. *Global livestock production systems*. Rome: Food and Agriculture Organization of the United Nations (FAO) and International Livestock Research Institute (ILRI).

Robinson, T. & Pozzi, F. 2008. *Accessibility mapping in the horn of Africa: Applications for livestock policy*. IGAD Livestock Policy Initiative Working Paper number 45. Rome: Food and Agriculture Organization of the United Nations (FAO).

Robinson, T., Emwanu, T. & Rogers, D. 2007. Environmental approaches to poverty mapping: An example from Uganda. *Information Development* 23**:** 205-215.

Rogers, D., Emwanu, T. & Robinson, T. 2006. *Poverty mapping in Uganda: An analysis using remotely sensed and other environmental data*. Pro-Poor Livestock Policy Initiative (PPLPI) Working Paper number 60. Rome: Food and Agriculture Organization of the United Nations (FAO).

Rogers, D., Wint, G.R.W. Alexander, N., Pozzi, F. & Robinson, T.P. 2011. *Wealth Index mapping in the Horn of Africa*. Animal Production and Health Division Working Paper number 4. Rome: Food and Agriculture Organization of the United Nations.

Rogers, D.J. 1997. *Satellite Imagery and the prediction of tsetse distributions in East Africa. Diagnosis and Control of Livestock Diseases using Nuclear and related techniques*. Vienna: International Atomic Energy Agency.

Rogers, D.J. 2000. Satellites, space, time and the African trypanosomiases. In: Hay, S.I., Randolph, S.E. and Rogers, D.J. (Eds.) *Remote Sensing and Geographic Information Systems in Epidemiology*, pp. 129-171. London: Academic Press.

Rogers, D.J. & Williams, B.G. 1994. *Tsetse distribution in Africa: seeing the wood and the trees*. Large-scale ecology and conservation biology. 35th symposium of the British Ecological Society with the Society for Conservation Biology, (1993). University of Southampton, U.K., Blackwell Scientific Publications.

Rogers, D.J., Hay, S.I. & Packer, M.J. 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Annals of Tropical Medicine and Parasitology* 90: 225-241.

Sammon, J.W. 1969. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers* 18: 401-409

Statistics South Africa. 2000. *Measuring poverty in South Africa*. Pretoria: Statistics South Africa. 107 pp.

Thornton, P.K, Kruska, R.L., Henninger, N., Kristjanson, P.M., Reid, R.S., Atieno, F., Odero, A. & Ndegwa, T. 2002. *Mapping poverty and livestock in the developing world*. Nairobi: International Livestock Research Institute (ILRI). 124 pp.

Thornton, P.K., Jones, P.G., Owiyo, T., Kruska, R.L., Herrero, M., Orindi, V., Bhadwal, S., Kristjanson, P., Notenbaert, A., Bekele, N. & Omolo, A. 2008. Climate change and poverty in Africa: Mapping hotspots of vulnerability. *African Journal of Agricultural and Resource Economics* 2: 24-44.

UBOS. 2002. *Uganda Population and Housing Census*. Kampala: Uganda Bureau of Statistics (UBOS).

UBOS. 2003. *National Household Survey 2002/2003. Report on the Socio-economic survey*. Kampala: Uganda Bureau of Statistics (UBOS).

UBOS. 2009. *Statistical Abstract*. Kampala: Uganda Bureau of Statistics (UBOS).

World Bank. 2000. *World development report 2000/2001: Attacking poverty*. New York: Oxford University Press.

ONLINE PUBLICATION SERIES

FAO ANIMAL PRODUCTION AND HEALTH WORKING PAPER

1. The use of cash transfers in livestock emergencies and their incorporation into Livestock Emergency Guidelines and Standards (LEGS), 2011 (E)
   http://www.fao.org/docrep/014/i2256e/i2256e00.pdf
2. Mapping supply and demand for animal-source foods to 2030, 2011 (E)
   http://www.fao.org/docrep/014/i2425e/i2425e00.pdf
3. Notes on Livestock, Food Security and Gender Equity, 2011 (E)
   http://www.fao.org/docrep/014/i2426e/i2426e00.pdf
4. Wealth Index mapping in the Horn of Africa, 2011 (E)
   http://www.fao.org/docrep/014/i2427e/i2427e00.pdf
5. Evolution du secteur avicole en Tunisie, 2011 (F)
   http://www.fao.org/docrep/015/i2549f/i2549f.pdf
6. Status of animal nutrition research and development activities in Tajikistan, Kyrgyzstan and Azerbaijan, 2012 (E)
   http://www.fao.org/docrep/015/i2582e/i2582e00.pdf
7. An assessment of the socio-economic impacts of global rinderpest eradication – Methodological issues and applications to rinderpest control programmes in Chad and India, 2012 (E)
   http://www.fao.org/docrep/015/i2584e/i2584e00.pdf
8. Use of lesser-known plants and plant parts as animal feed resources in tropical regions, 2012 (E)
   http://www.fao.org/docrep/015/i2629e/i2629e00.pdf
9. Poverty mapping in Uganda – Extrapolating household expenditure data using environmental data and regression techniques, 2012 (E)

Availability: April 2012

E - English
F - French
** In preparation