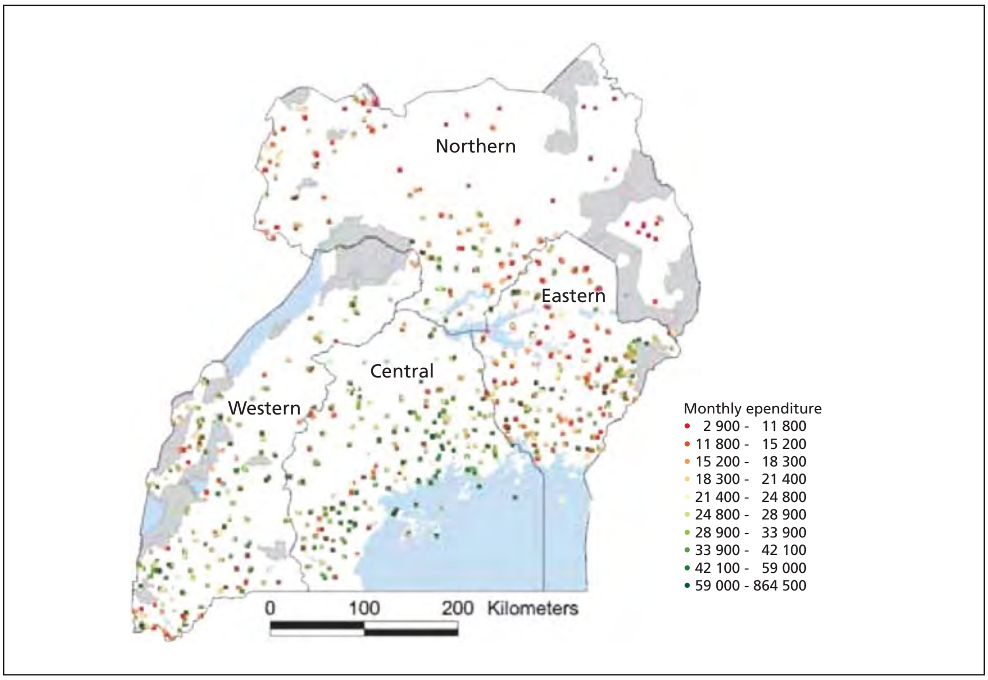
Data

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:

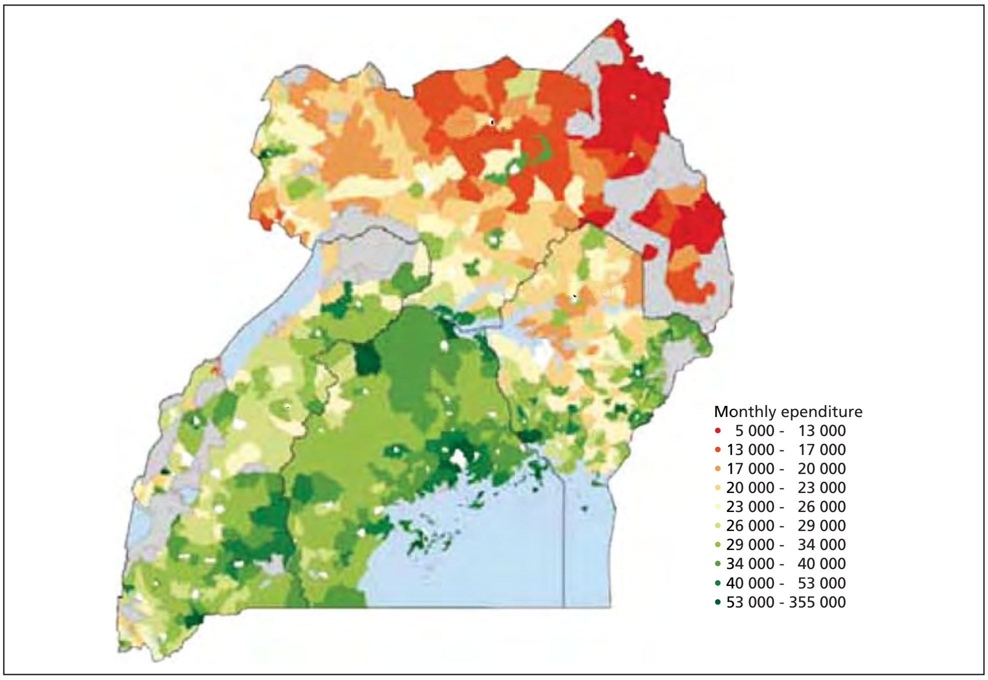
$$y_i = A_i'\beta_i + \varepsilon_i$$

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

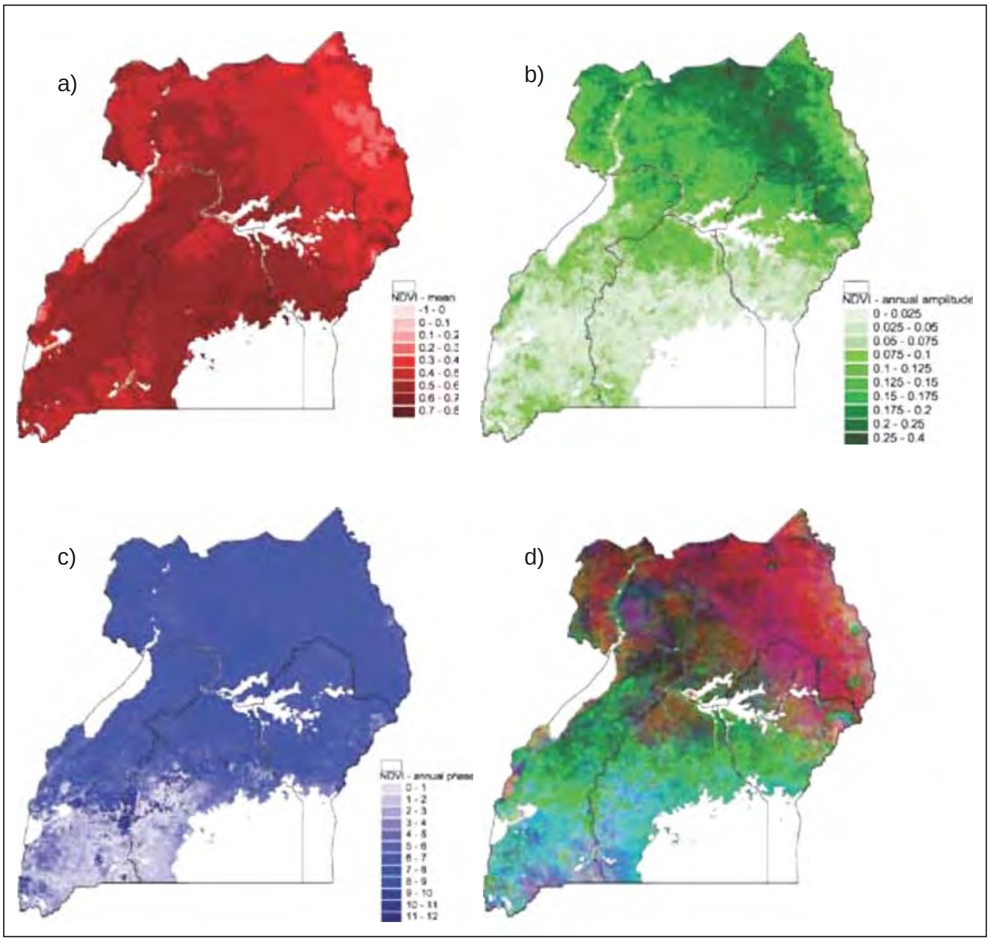
The majority of the explanatory variables used in the regression modelling were satellite-derived, and most came from the 1km global Advanced Very High Resolution Radiometer (AVHRR) dataset made available by the National Aeronautics and Space Administration (NASA) Pathfinder program. These data were 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).

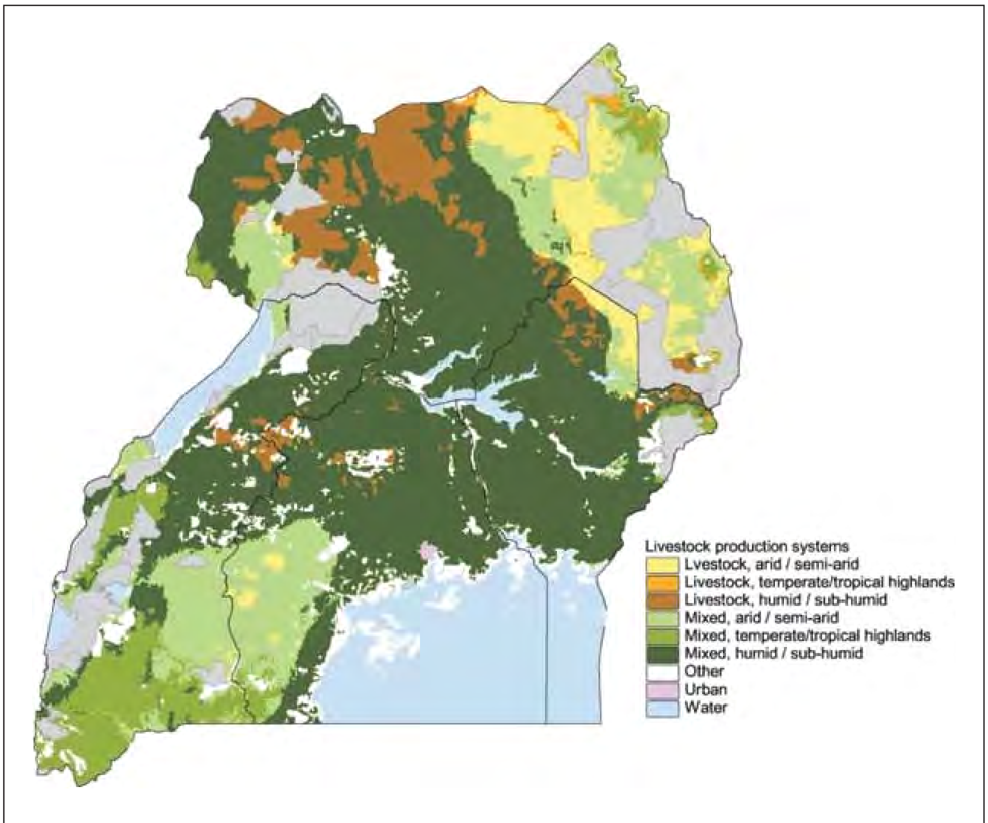
The classification is based on four modes of production: livestock grazing; rain-fed mixed crop and livestock production; irrigated mixed crop and livestock production; and landless livestock production. These are further split among three agro-ecological zones defined by LGP and temperature: arid and 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.