ISSN 2221-8793

4

FAO ANIMAL PRODUCTION AND HEALTH



working paper

WEALTH INDEX MAPPING IN THE HORN OF AFRICA

Pro-Poor Livestock Policy Initiative

A Living from Livestock

FIAT PANIS

Cover photographs:

Left: FAO/Jose Cendon
Centre images: FAO/Giulio Napolitano
Right image: FAO/Sarah Elliott

4

FAO ANIMAL PRODUCTION AND HEALTH
working paper

WEALTH INDEX MAPPING IN THE HORN OF AFRICA

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 2011

Recommended Citation
FAO. 2011. *Wealth Index mapping in the Horn of Africa*, by D.J. Rogers, G.R.W. Wint, N. Alexander, F. Pozzi & T.P. Robinson. Animal Production and Health Working Paper. No. 4. Rome.

Keywords
Horn of Africa; Wealth Index; Environmental poverty mapping; livestock; discriminant analysis; multi-temporal, Fourier-processed satellite data.

Authors' details
David Rogers is Professor of Ecology at the University of Oxford, and heads the Trypanosomiasis and Land use in Africa (TALA) Research Group within the Spatial Ecology and Epidemiology Group in the Department of Zoology. The Fourier processing and discriminant analytical techniques used here have been developed by the TALA research group.

William Wint is Managing Director of the Environmental Research Group Oxford (ERGO) and a Senior Research Associate of the Zoology Department at Oxford University. For the past 20 years he has focussed on developing and using new techniques of geographical analysis and mapping of agriculture and the environment, animal and human disease risks, and climate change.

Neil Alexander works for the ERGO, with a focus on the manipulation of large environmental and remotely sensed datasets in preparation for use in statistical models, especially related to disease risk.

Francesca Pozzi conducted this work as a consultant to the IGAD LPI. Her research interests focus on the use of geographic information systems (GIS), remote sensing and statistical tools to analyze poverty distribution and the linkages between poverty and livestock, agriculture, population distribution and 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 978-92-5-107043-7 (PDF)

Table of contents

List of tables and figures

TABLES

FIGURES

Acknowledgments

The authors would like to acknowledge Mark Montgomery for his inputs on the DHS Wealth Index, and David Benz of the TALA Research Group for temporal Fourier image processing of the MODIS satellite data used in this analysis.

Francesca Pozzi, Claudia Ciarlantini and Carmen Hopmans are acknowledged for the design, layout and formatting of the working paper.

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 into a single series 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.

This paper follows on from a previous FAO study that used remotely sensed and other environmental data to map poverty in Uganda (FAO, 2006) and extends it to the Horn of Africa, incorporating additional environmental and sociological variables. Furthermore, instead of using a direct measure of poverty, this study investigates the use of the Demographic and Health Survey (DHS) Wealth Index (WI) as a proxy for a regional welfare measure.

Abbreviations

AGAL	FAO Livestock Information, Sector Analysis and Policy Branch
AVHRR	Advanced Very High Resolution Radiometer
CIAT	Centro International de Agricultura Tropical
CIESIN	Center for International Earth Science Information Network
DCW	Digital Chart of the World
DHS	Demographic and Health Survey
DMSP	Defence Meteorological Satellite Program of the United States
EARS-NL	A high-tech remote sensing company, based in the Netherlands
EC	European Commission
ERGO	Environmental Research Group Oxford
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organisation
FAO-FSNAU	FAO Food Security and Nutrition Analysis Unit - Somalia
GIS	Geographic Information System
GLW	Gridded Livestock of the World
GPS	Global Positioning System
GPW	Gridded Population of the World
GRUMP	Global Rural and Urban Mapping Project
HF	Human Footprint
HII	Human Influence Index
IGAD	Inter-Governmental Authority on Development
IGAD LPI	IGAD Livestock Policy Initiative
ILRI	International Livestock Research Institute
IRD	Institute for Resource Development
JRC	Joint Research Center of the European Commission
LST	Land Surface Temperature
MIR	Middle Infra-Red
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalised Difference Vegetation Index
NGO	Non-Government Organisation
NIMA	US National Imagery and Mapping Agency (former name of the National Geospatial-Intelligence Agency)
NOAA	National Oceanic and Atmospheric Administration
PCA	Principal Component Analysis
PAAT	Programme Against African Trypanosomiases
PPLPI	Pro-Poor Livestock Policy Initiative
SAE	Small Area Estimate
TALA	Trypanosomaiasis And Land-use in Africa
USAID	United States Agency for International Development
WCS	Wildlife Conservation Society
WI	Wealth Index

Executive summary

Poverty measures are usually based on economic indicators, such as income or expenditure, or on a number of social indicators such as life expectancy, under-five mortality, nutritional status, and so on, usually collected through household surveys. Recently, researchers and policy makers started to analyze poverty through the use of geographically disaggregated indicators that provide information about the spatial distribution of inequality and poverty within a country: these are usually referred to as 'poverty maps'. The most common approach to poverty mapping is the small area estimation technique, developed by the Word Bank, which combines census and survey data to provide sub-national estimates of welfare.

Another, more recent approach involves the combination of household survey data with a suite of environmental and other spatial variables not only to map but also to try and explain and possibly predict the distribution of poverty. In Uganda, satellite data have proved useful in understanding and possibly predicting the causes of poverty. Such imagery, when appropriately processed, captures habitat seasonality associated with the growing seasons for crops, or transmission seasons for vector-borne and other diseases. These seasonal signals were used within a discriminant analytical framework to describe the different levels of household expenditure.

By using an appropriately reconstructed Wealth Index (WI) from the Demographic and Health Survey (DHS) data as a proxy for poverty, the same approach can be extended to larger regions. The results of the present study conducted in the Horn of Africa indicate that the lowest levels of Wealth Index are associated with dry conditions; intermediate Wealth Index levels are associated with moister, greener conditions; and high Wealth Index levels are associated with less green conditions and human activity - high population densities and proximity to population centres. The detailed nature of the relationships remains to be investigated.

The approach described here shows that it is possible to use the DHS WI as a regional poverty indicator, provided that it is reconstructed from a set of common indicators from the individual, national DHS surveys. Many questions remain to be addressed in developing the environmental approach to poverty mapping, but the present analysis confirms that environmental variables are important correlates of human welfare and may be used to describe welfare levels across climatically and sociologically diverse regions.

Introduction

FAO (2006) showed how the spatial pattern of household expenditure in Uganda could be described in terms of environmental data derived from satellites. This approach to poverty mapping escapes from the somewhat circular approach adopted by the small area estimation technique (Hentschel *et al.*, 2000; Elbers and Lanjouw, 2000; World Bank, 2000) that exploits the internal correlations within socio-economic data sets. Poorer, in contrast to richer, people cannot afford bicycles or radios, and are unlikely to have access to clean drinking water. One might therefore use the possession of such assets, or lack thereof, to describe any chosen single index of poverty, such as household expenditure. Linking the same index of poverty to environmental data begins to break out of this circularity and looks for causes of poverty rather than the consequences of it. The underlying assumption in this approach is that people in rural settings are poor because their environments fail to provide the goods and services available to richer people. Soil fertility, good health, access to fuel and water all have environmental correlates for which satellite data may act as proxies: people are often poor because of an inadequate supply of these vital resources. By incorporating the driving factors that are associated with the different levels of poverty, the modelling approach allows not only for a description, but potentially also for an explanation and, ultimately, a prediction of the distribution of poverty.

It is obvious that a strict environmental approach to poverty mapping will apply best to subsistence agricultural systems, where external inputs in the form of soil improvements (e.g. fertilisers), carbon subsidies (e.g. oil for tractors) or cash subsidies (e.g. tariffs) are minimal or lacking. It cannot apply also to urban communities, which are variously connected to external cash economies, and so less dependent on the immediate environment.

The vast majority of rural people in less developed countries, and especially in sub-Saharan Africa, still practice subsistence agriculture, where environmental constraints are likely to be critical and limiting to welfare. This assumption appeared to be borne out by the Uganda analysis that used a set of socio-economic data from the Uganda Bureau of Statistics and a set of environmental variables, including satellite data derived from the National Oceanic and Atmospheric Administration (NOAA) satellites' Advance Very High Resolution Radiometer (AVHRR). This satellite series provides a more or less uninterrupted sequence of monthly global imagery from the early 1980's to the late 1990's. Such imagery, when appropriately processed, captures habitat seasonality associated with the growing seasons for crops, or transmission seasons for vector-borne and other diseases, and these seasonal signals were used within a discriminant analytical framework (that naturally allows for any non-linearity in the relationship between the index of poverty and the environmental data) to describe the different levels of household expenditure. The analysis showed how the correlations between satellite data and household expenditure increased in strength from finer to coarser spatial resolutions (a common feature of all poverty mapping exercises) and were equivalent to, or better than the small area mapping results at a spatial resolution of *c.* 20 to 30 km (FAO, 2006; Robinson *et al.*, 2007).

Following on from the Uganda study, the Inter-Governmental Authority on Development (IGAD) supported an initiative that sought to extend this approach to the Horn of Africa, including Sudan, Eritrea[1], Ethiopia, Djibouti, Somalia, Kenya and Uganda. This working paper describes the results of this new study undertaken to support the EC funded IGAD Livestock Policy Initiative (LPI). All data and results are archived in digital format (compatible with ESRI GIS software) on the IGAD Livestock Information Portal: http://www.igad-data.org/.

The next section discusses the analytical methodology and the data used. This includes a description of the DHS Wealth Index (WI) and the steps required to construct a Regional WI. The predictor variables are then described as is the modelling approach, which involves non-linear discriminant analysis applied to the socioeconomic and environmental data. The following section presents the results of the analyses – exploring issues of data aggregation by comparing the analysis of clusters of household data against that based on individual households. In the final section we draw some conclusions from our analyses and discuss some ways in which the environmental approach to regional poverty mapping may be taken forward in the future.

[1] Eritrea has currently suspended its membership of IGAD.

Methods

In this section we describe the input data and the methodology used to map the Wealth Index in the Horn of Africa.

SOCIO-ECONOMIC SURVEYS

For many economists, household income would be the indicator of choice to determine economic status. It is, however, extremely difficult to measure income accurately for a number of reasons. People often try to hide their income from interviewers, for example by not providing accurate estimates; all elements of income may not be shared among household members; or income may vary considerably depending on the time of year. As well as difficulties in estimating income, it may not represent an equivalent estimate of welfare across different social contexts. For example, in pastoralist societies welfare tends to be more closely related to livestock assets held rather than to income generated. An alternative approach is to measure household consumption expenditure. Consumption expenditure estimates are generally easier to collect and more readily standardised across countries (World Bank, 2003). Moreover, consumption is thought to be a more stable measure of poverty over time than is income in agricultural economies (Deaton and Zaidi, 2002).

Income or expenditure data are collected either through household surveys that are specifically designed to collect such information (welfare monitoring surveys) or through more generic surveys, primarily designed to collect and update social and demographic indicators. Such surveys may also include socio-economic modules in their questionnaires.

In most IGAD member states, the World Bank, in collaboration with the national governments or other international agencies, has conducted a series of household surveys to collect socio-economic data at the household level. In some cases, such as Uganda, the central government conducts regular national household surveys, with very similar objectives. These surveys usually contain information on demographics, health, education, employment, income and expenditure, as well as household characteristics and agricultural and livestock assets. Whist these surveys provide good relative comparisons of welfare within a country at a particular time, the results are not comparable across countries (and not necessarily across different time periods in a particular country). This means that regional analyses and comparisons are not possible using such data.

Another type of household survey that may be able to overcome these problems of standardisation is the Demographic and Health Survey (DHS). The DHS program was established by the United States Agency for International Development (USAID) in 1984. It was designed as a follow-up to the World Fertility Survey and the Contraceptive Prevalence Survey projects. The DHS project was established at the Institute for Resource Development, Inc. (IRD), which was subsequently acquired in 1989 by Macro International Inc. (OCR Macro), the company that manages the collection, analysis and dissemination of data, and has been implemented in overlapping five-year phases. In 1993 DHS was folded into USAID's multi-project MEASURE program as MEASURE DHS+, which incorporated traditional DHS features, expanded the content on maternal and child health, and added biomarker testing to numerous surveys. The MEASURE DHS program is still funded principally by USAID with contributions from other donors.

The objectives of the DHS program are, among others, to provide decision-makers in participating countries with improved information and analyses in support of making informed policy choices; to improve coordination and partnerships in data collection at the international and country levels; and to develop in participating countries the skills and resources necessary to conduct high-quality demographic and health surveys. The basic approach of the DHS program is to collect data that are comparable across countries. To this end, standard model questionnaires have been developed, accompanied by user guides and manuals. Since 1984, more than 130 nationally representative household surveys in about 70 countries have been completed under the DHS project. Many of the countries have conducted multiple DHS surveys to establish trends, enabling them to gauge progress in their programs.

The DHS surveys are designed to collect household data on marriage, fertility, family planning, reproductive health, child health and HIV/AIDS (Rutstein and Rojas, 2003). They do not collect information on economic measures of poverty, such as income or expenditure, but data are collected about the dwelling itself, such as the source of water, type of toilet facilities, materials used to construct the house and ownership of various assets. These asset indices may be used as a proxy for the wealth status of the household (see the sub-section below on the DHS Wealth Index).

The most recent DHS survey data are accompanied by global positioning system (GPS) coordinates at the cluster level, where a cluster is usually a census enumeration area, sometimes a village in rural areas or a city block in urban areas. Collecting only one location point for a cluster greatly reduces the chance of compromising the confidentiality of respondents, but it is enough to allow the integration of multiple datasets for further analysis (Montana and Spencer, 2004).

DHS surveys have been carried out in all IGAD member states with the exception of Djibouti and Somalia. In Sudan the surveys are representative only of large administrative units so, for the current analysis, we used only the datasets for Eritrea, Ethiopia, Kenya and Uganda to develop welfare models, though predictions were made for all countries in the region. For each country, the most recent dataset available was used, specifically: Eritrea 2002, Ethiopia 2005, Kenya 2003 and Uganda 2001[2]. These four surveys included a total of 37 352 households, which were grouped into 1 519 geo-located clusters.

The DHS Wealth Index

Whilst the DHS surveys do not collect information on income or expenditure, a proxy that can be used is the Wealth Index (WI), which is constructed from a number of indicators that are thought to be correlated with a household's economic status (Rutstein and Johnson, 2004). Component indicators include, for example, possession of assets such as a television, radio, telephone or refrigerator, and variables describing the dwelling, such as the type of flooring, water supply, sanitation facilities and number of people per sleeping room.

[2] More recent datasets were subsequently released for Kenya and Uganda (2008/2009 and 2006 respectively) but they were not available at the time this analysis was conducted.

The WI, as computed from individual national surveys, cannot be used for direct cross-country comparisons since the indicators included vary from country to country. The WI is in fact a relative measure of wealth within a given survey (Rutstein and Johnson, 2004). In an FAO study, the authors discuss the WI in relation to other welfare estimates within individual countries of the IGAD region, showing a good correlation between the different measures (FAO, 2008). The objective here is to explore its value as a regionally consistent measure of welfare that can be used to produce regional poverty maps.

The WI is constructed by way of a Principal Component Analysis (PCA) on the recorded set of assets and services (Filmer and Pritchett, 2001). DHS uses the SPSS factor analysis procedure (see for example Field, 2005). This procedure first standardises the indicator variables (by calculating z-scores); then the factor coefficient scores (factor loadings) are calculated; and finally, for each household, the indicator values are multiplied by the loadings and summed to produce final values on each PCA axis. Each resulting sum is a standardised score with a mean of zero and a standard deviation of one. In the present analysis, following convention, only the first of the factors produced is used to represent the WI.

Many of the indices of poverty cannot easily be put on a quantitative scale (Rutstein and Johnson, 2004), but they can usually be coded in some way, and hence included in quantitative analyses. PCA uses only quantitative data, but this can include binary or dummy-coded qualitative data such as the presence or absence of something. Care must be taken, however, not to introduce variables falsely giving the appearance of a quantitative scale. For example, one could assign scores of '1' to the possession of a bicycle, '3' for a motorcycle and '5' for a car, but such a weighting scheme would be arbitrary and would not provide an acceptable pre-treatment of data destined for PCA. Dummy coding these same variables (i.e. creating a separate variable for each mode of transport and assigning scores of 1 or 0 to indicate presence or absence, respectively) would, however, be acceptable.

The use of a single score (PCA axis 1) for any index of wealth assumes that the majority of the variation within the dataset can be captured within this one dimension alone. Whilst PCA axis 1 by definition captures the largest percentage of the variance within the dataset, in complex data sets (such as those contributing to the WI) this may in fact be a small proportion of the total variance. There are as many axes within a PCA as there are variables (n) in the original data set, and since each axis captures some of the variance (in decreasing amounts from PCA axis 1 to PCA axis n). The larger the number of variables, the less likely it is that PCA axis 1 captures an absolute majority of the total variance (in fact it will only do so if all the indicator variables are highly correlated with each other; in which case some of these variables are redundant, and could be excluded from the questionnaires, thus saving time and resources). Using PCA axis 1 scores alone to capture poverty must therefore be approached with caution.

As mentioned above, each of the country-specific datasets has in the past been subjected to a separate PCA. This has two consequences. First, each country's PCA will be derived from a different set of input variables, only some of which might be shared with other countries. Second, even if the same set of variables were used for each country (with PCA again carried out separately for each country), the WI cannot be directly compared between countries. This is because all PCA scores are specific to the datasets being analysed, and the PCA's outputs are mean-centred scores on each of the PCA axes. To illustrate this further, consider two countries, A and B, one on average much richer than the other: there is a range of variation of wealth around each country's average wealth. Let the richest people in the poorer country be poorer than the poorest people in the richer country (i.e. there is no overlap in the wealth of any of the citizens in the two countries). PCAs carried out on each country's data will put the relatively wealthier people of each country on the positive side of its PCA axis 1[3], and the relatively poorer people on the negative side. The WI values of certain individuals within both countries may therefore be the same, despite the fact that they are not equally wealthy in absolute terms. They are only equally wealthy in relative terms, and compared only with their fellow country-men and women, not with the foreigners from the other country. Individual-country PCAs thus hide the difference in mean wealth between the two countries.

Construction of a Regional Wealth Index

The solution to this dilemma, of course, is to carry out a single PCA for all countries together. To do this we must use a set of input socio-economic variables common to all countries. Table 1shows the DHS indicators used to construct the WI in individual countries in the Horn of Africa (Eritrea, Ethiopia, Kenya and Uganda), and whether they were used in the present analysis to build the Regional WI. Some of the indicators were excluded because data were not available in all 4 surveys.

It is also important to check that combining countries' data in this way does not distort the results from the individual countries' PCA (it should not do so because the data themselves are not transformed in any way within PCA: each point stays at the same distance from all other points in the dataset within the rotated axes as it was within the un-rotated axes). We want to achieve both the same relative ranking of individuals within each country as was obtained by the country-specific PCA and also a single-scale measure of WI applicable across all countries combined. In this way, the absolute richest individuals across all countries will end up with the same WI scores; similarly for people at all other absolute levels of wealth or poverty[4].

In order to check that a regional measure of WI was accurately reflecting the previous within-country estimates, we first calculated the correlations between the regional and country-specific WIs (shown in Tables A2 to A5 in Annex 1). The latter were calculated in two ways: first using all of the variables available within that country and secondly using only those variables common to all countries. It was expected, and generally found, that correlations of Regional WI with the first sort of within-country WI were less strong than they were with the second (respectively the red and blue highlighted figures in Tables A2 to A5), but the differences were small.

[3] PCA axis scores may have reversed signs, with the richest people ending up with the highest negative scores, and poorest people with the highest positive scores. It is the absolute difference between scores that is the real index of absolute differences in the WI.

[4] Combining data across countries in this way will give the same relative weight within the PCA – and therefore the calculation of WI - to the possession, for example, of a bicycle in all countries. This may not actually reflect reality on the ground. The song "Oh Lord, won't you buy me a Mercedes Benz?" might change in the poorest country of all to one that simply requests the Almighty's supply of a decent bicycle.

Table 1. Field survey indicator variables.

Indicator	Used to compute Regional WI
Has electricity	Y
Has radio	Y
Has television	Y
Has refrigerator	Y
Has bicycle	Y
Has motorcycle	Y
Has car	Y
Has telephone	Y
Drinking water is piped in residence	Y
Drinking water is piped in public tap	Y
Drinking water from well in residence	Y
Drinking water from public well	Y
Drinking water is from surface water	Y
Drinking water is rainwater	N
Other source of drinking water	Y
Has own flush toilet	Y
Uses shared flush toilet	Y
Has pit latrine	Y
Has ventilated pit latrine	Y
Uses bush as latrine	Y
Uses other type of latrine	Y
Has dirt, earth principal floor in dwelling	Y
Has wood planks principal floor in dwelling	Y
Has tile flooring	Y
Has cement flooring	Y
Has other type of flooring	Y
Has natural material roofing	N
Has corrugate iron roofing	N
Has roofing tiles	N
Has other roofing	N
Number of members per sleeping room	N
Has domestic servant	N
Household works own or family ag. land	N

Secondly, in order to ensure compatibility across datasets we also examined the weightings of individual indicators of wealth[5] across the within-country analyses. This addresses questions of the following sort (and raised in footnote 4) 'Does a bicycle in Eritrea have the same PCA loading as a bicycle in Uganda?' In general the correlations were very good (Figure1). The exceptions to this general rule were

[5] Weightings in PCA are the natural cosines of the angle between the original (raw variable) axis and each rotated (PCA) axis. The axis of a raw variable that is highly correlated with the WI will have a small angle to the rotated axis, and will therefore end up with a high weighting (since COS(0 degrees) = 1, and COS (90 degrees) = 0).

also of interest. For example, the variable *WCBush* (indicating use of surrounding vegetation for toilet purposes) had a weighting of -0.15 in Eritrea but only of -0.06 in Uganda. This indicates a greater (negative) correlation of this variable with the WI in Eritrea than in Uganda. Generally differences were smaller than this.

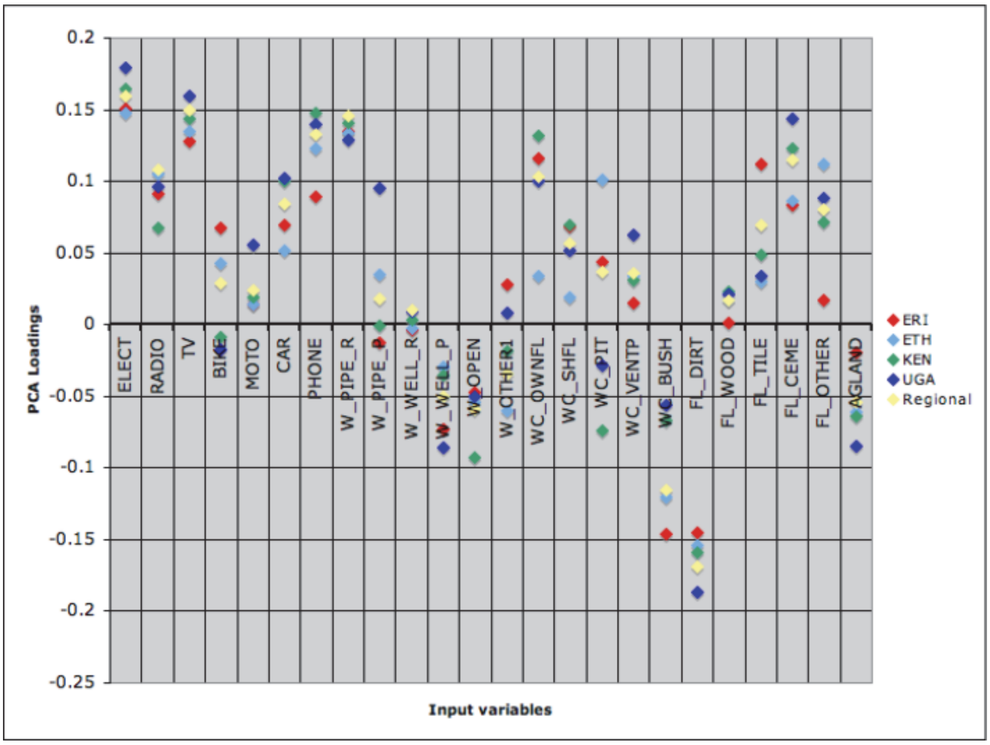
After the above comparisons were examined, it was decided that a single Regional WI provides an acceptable measure of region-wide, rather than country-specific, welfare, and all the results here are based on the Regional WI values, using the common variables listed in Table 1. Figure 2 shows the resulting Regional WI for the geo-referenced clusters.

PCA may be carried out on the raw data or on the standardised data (in the latter case the mean value for that variable is subtracted from the data value and the result is divided by the standard deviation of the variable concerned; standardised variables tend to be in the range -3.0 to +3.0). The original WI analyses first standardised the input socio-economic data (Rutstein and Johnson, 2004), and that was also the practice adopted here.

The WI values were either based on the data aggregated to the cluster level (i.e. the values of each socio-economic variable were the average for all households within each cluster), or on the individual household survey data. There were 37 352 households in the entire dataset, grouped into 1 519 geo-located clusters. Both sets of data were eventually modelled.

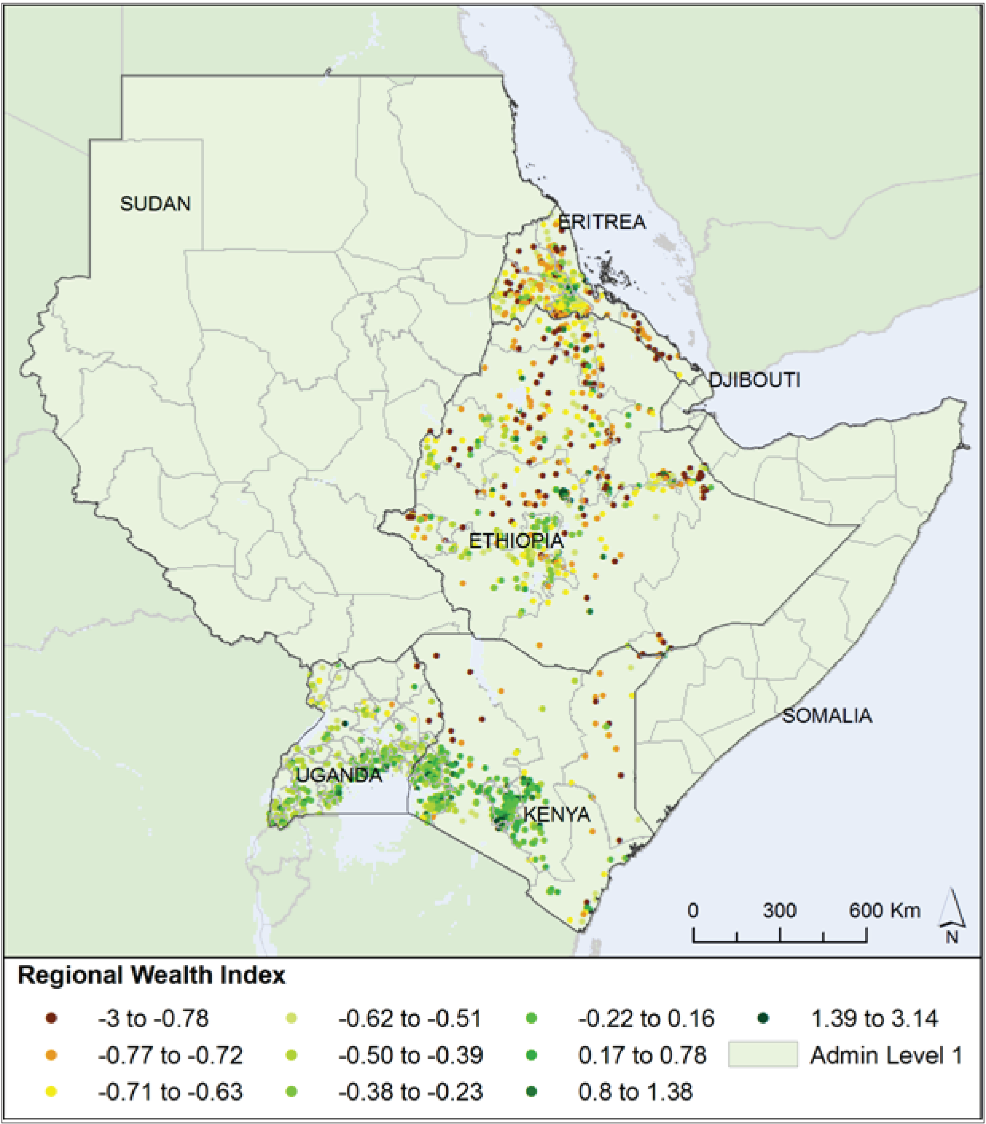
A first test was carried out to investigate whether the authors could repeat the results of the original analyses for each country. Care was taken to establish that we knew exactly how the data had been processed within the PCAs carried out previously. We were able to obtain the same values for the WI as the original survey analyses reported.

Figure 1. PCA Loadings.



Note: The figure shows PCA loadings (y-axis) of the input variables (x-axis) for each country separately, and regional loadings for the same variables. The fact that the individual country values tend to share similar PCA loadings indicates a similar contribution of each variable to the country-specific WI calculations. The fact that the regional loadings are within the range of values of the individual countries indicates that calculating a regional rather than country-specific WI does not change the relative contribution of each variable to the single, regional index.

Figure 2. Regional Wealth Index, for the 1 519 geo-referenced DHS clusters.



REMOTELY SENSED ENVIRONMENTAL DATA

Since the original environmental analysis of poverty in Uganda was conducted (FAO, 2006; Robinson *et al.*, 2007), the team in Oxford has processed the 2001 to 2005 series of satellite data from the MODerate-resolution Imaging Spectroradiometer (MODIS) sensor on board the newer Terra and Aqua satellites. These data are spectrally similar to (though not identical with) the AVHRR channels, used in the Uganda study, but offer much better geo-registration and spectral stability. In short, they are a better measure of environmental conditions for the period in question. The MODIS datasets used in this analysis include daytime and night-time land surface temperature (LST), the Middle-infrared (MIR) reflectance and the vegetation indices: Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI)[6], and an Evapotranspiration layer, at a coarser resolution.

[6] NDVI is calculated as [Near infrared (NIR) – RED] / [NIR + RED], where NIR is MODIS band 2 and RED is MODIS band 1. EVI is calculated as (2.5 * [NIR – RED] / [NIR + 6*RED – 7.5*BLUE + 1], where BLUE is MODIS band 3.

In addition to the MODIS data, the team obtained actual and potential evapotranspiration and precipitation data derived from METEOSAT, provided by EARS-NL, a high-tech remote sensing company, based in the Netherlands.

Both the MODIS and the METEOSAT data (listed in Table 2), were temporal Fourier processed to extract the seasonal fingerprint of each pixel in each channel (Scharlemann *et al.*, 2008). Temporal Fourier analysis transforms time-series satellite observations into a set of (uncorrelated) sine curves, or harmonics, of different frequencies, amplitude and phases that often have a clear biological interpretation (Rogers and Williams, 1994). For each variable the Fourier process outputs the mean, amplitude and phases of annual, bi-annual and tri-annual cycles, and, in addition, the minimum, maximum and variance of the smoothed data values.

ADDITIONAL DESCRIPTOR DATA

Additional data made available to the models are also listed in Table 2. These included distance to markets, population and livestock densities and the probability of occurrence of different tsetse species. Some of these variables (e.g. livestock densities) were themselves derived by modelling and the process by which they were derived is explained in Annex 2.

The population density layers used in this analysis were those developed by CIESIN, and in particular the Gridded Population of the World (GPW) version 3, (CIESIN and CIAT, 2005) and the Global Rural and Urban Mapping Project (GRUMP) (CIESIN *et al.*, 2004). Both GPW and GRUMP gridded data are derived from a simple proportional allocation gridding algorithm of national and sub-national level population data. GPW data are available at a resolution of 2.5 arc-minutes (Balk and Yetman, 2004). GRUMP distinguishes urban and rural population from around the year 2000 and is available at the finer spatial resolution of 30 arc-seconds (Balk *et al.*, 2004). GRUMP also supplies a database of human settlements, which comprises some 55 000 cities and towns with populations of 1 000 or more, and a map of urban extents, which was derived largely from the night-time lights (Elvidge *et al.*, 1997).

In order to determine the contribution of human population, not only in terms of population density, but also in relation to its impact on the environment, it was decided to include also the Human Footprint layer, from the Last of the Wild project (WCS and CIESIN, 2002; Sanderson *et al.*, 2002). The Human Footprint (HF) layer is produced through an overlay of a number of global data layers that represent the location of various factors presumed to exert an influence on ecosystems: human population distribution, urban areas, roads, navigable rivers, and various agricultural land uses. The combined influence of these factors yields the Human Influence Index. The Human Influence Index (HII), in turn, is normalised by global biomes to create the HF data set, according to the methodology developed by Sanderson *et al.* (2002). HF values range from 1 to 100. A score of 1 in moist tropical forests indicates that that grid cell is part of the 1 percent least influenced or 'wildest' area in its biome, the same as a score of 1 in temperate broadleaf forests (although the absolute amount of influence in those two places may be quite different). The areas that have the least influence (HF grid values less than or equal to 10) are included in The Last of the Wild data set (WCS and CIESIN, 2002). For this analysis, version 1 was used.

Arguably some of these additional variables are effectively the same, but it was decided to include them all and see which ones the models selected. Variables which are perfectly correlated with those already selected will not themselves be selected in the step-wise approach adopted here, since the inclusion of effectively the 'same' data for a second time cannot possibly improve the fit of any model. The same argument applies to closely correlated variables; these too are unlikely to be selected together within the final predictor variable set unless some important differences between them allow an improvement in the overall model fit.

Table 2. Predictor variables used in the WI analysis. The 'total number of files' column indicates the number of files contained in each set of variables, which, in the case of the satellite data, results from the temporal Fourier processing. The last column indicates the number of files actually used in the model.

Data	Total no. of files	No. of files used in model	Resolution of original data	Source of original data
Vegetation Indices (NDVI, EVI)	102	60	1 km	NASA, MODIS version 4
Daytime and Night-time Land Surface Temperature (LST)				
Middle Infra-Red (MIR)				
Evapotranspiration			5 km	
Potential Evapotranspiration	10	10	3 km	EARS-NL
Actual Evapotranspiration				
Precipitation				
Global Land Cover	1	1	1 km	JRC, Global Land Cover 2000 (GLC2k)
Length of Growing period *	1	1	1 km	FAO/ILRI global livestock production systems
Digital Elevation Model (DEM)	1	1	1 km	NOAA, Global Land One-kilometer Base Elevation (GLOBE)
Slope	1	1		
Distance to Rivers	1	1	Calculated on 1 km grid	Local data where available, otherwise Africover (Eritrea and Kenya) and VMap0 (Djibouti)
Distance from Wetland	1	1	Calculated on 1 km grid	WWF, Global Land and Wetlands Database (GLWD)
Distance to Major roads	1	1	Calculated on 1 km grid	NIMA Digital Chart of the World (DCW) roads data layer, with the exception of Somalia, where the roads layer was provided by FAO-FSNAU
Distance to All Roads	1	1	Calculated on 1 km grid	Individual countries' road layers
Distance to Populated places (Gazetteer)	1	1	Calculated on 1 km grid	NIMA and GeoNames
Distance to Populated Places (Vmap0)	1	1	Calculated on 1 km grid	VMap Level 0

(cont.)

Table 2. *(cont.)*

Data	Total no. of files	No. of files used in model	Resolution of original data	Source of original data
Access to Markets *	1	1	Calculated on 1 km grid	CIESIN, Human settlements database from GRUMP, with the exception of Somalia, where market locations were provided by FAO-FSNAU
Population Density - GPW	1	1	5 km	CIESIN, Gridded Population of the World (GPWv3)
Population Density - GRUMP	1	1	1 km	CIESIN, Global Rural and Urban Mapping Project (GRUMP)
Urban Extents	1	1	1 km	CIESIN, Global Rural and Urban Mapping Project (GRUMP)
Human Footprint	1	1	1 km	WCS/CIESIN, Last of the Wild Project, v1
Night-time lights – City Lights	1	1	1 km	DMSP night-time lights
Night-time lights – Average Radiance	1	1	1 km	DMSP night-time lights
Cattle Density *	1	1	5 km	FAO, Gridded Livestock of the World (GLW)
Camel Density *	1	1		
Sheep Density *	1	1		
Goat Density *	1	1		
Pig Density *	1	1		
Chicken Density *	1	1		
Cropping *	1	1	1 km	
Tsetse *	3	3	1 km	FAO PAAT information system
IGAD Mask	1	1	1 km	Land/Water recode on NDVI image (MODIS)
Country Layer			Gridded at 1 km	FAO Global Administrative Unit Layers

Note: * indicates variables that were derived from interpolated or modelled data.

MODELLING APPROACH

As in the Uganda case study, the modelling approach was based on non-linear discriminant analysis (FAO, 2006), which allows the prediction not only of binary (presence/absence) data, but also of continuous (i.e. socio-economic) and multiple category data. In the present discriminant analyses the WI data (our 'poverty' measure), divided into ten approximately equal-sized categories, were the dependent variable and the environmental and temporal Fourier data layers were the independent or predictor variables.

The algorithm examined the predictor variables one at a time to discover which one maximised the discriminant criterion selected by the user (in our case kappa, the index of agreement between model-predicted and observed data). This variable became the first selected variable of the eventual predicted WI map. The algorithm then went through the remaining variables, again one at a time, to select which one, in association with the first one selected, maximised the same discriminant criterion. The algorithm continued in this stepwise fashion until a pre-set number of variables

(10 in the present case) was selected. The set of selected variables was then used to make a map of the model-predicted poverty categories. More details on the discriminant analytical methods and on various metrics of model accuracy are provided by in Annexes C and D of FAO (2006).

Two types of model were run. In the first, the clustered data were used, in the second the individual household data. By definition, the predictor variable values for all households within the same cluster must be the same (because the households are given the same geo-location), and each cluster of households formed a single point (i.e. a single mean WI) which could go into only one of the ten WI categories in the model based on the clustered data. For the model based on individual household data, however, each household in a cluster might be assigned to a different WI category, depending on its individual WI. Thus one might expect a different set of predictor variables for each WI category of the clustered or individual household data.

A first model was run using the clustered data (Model 1). Then a second model was run using the individual household data (Model 2). The mapped results of these two models differed in appearance, and it was thought that this might arise because the category boundaries in the two models differed (although, it seemed, only marginally). To test whether the differences were due just to category boundaries, a third model was run using the clustered data sorted into ten categories and using the same category boundaries as were used in the household level model (Model 3).

Results

The predictions of Models 1 to 3 are shown in Figures 3, 4 and 5 respectively, with further details in the Annex Tables A6, A7 and A8 respectively.

Key predictor variables for all three models involved human populations, either the human footprint layer (Models 1 and 2) or the GRUMP human population density surface (Model 3). The 'spaghetti-like' appearance of Model 3 (Figure 5) could be explained by the second selected variable (distance to roads), which the other models do not have. In all cases, the WI increases with increasing values of the human population variables. Because the same variable was chosen first by the first two models, the subsequent variables are quite similar. The choice of GRUMP as the first variable in Model 3 is likely to have affected the choice of all subsequent variables in Model 3 that appear to be quite different from the variables chosen in the first two models. This is most likely due to the correlation structure of the data: it is possible that the sets of variables chosen by Models 1 and 2 are relatively closely correlated with the variables chosen by Model 3.

The general consensus on these three Models is that Model 2 is the 'best' in capturing what we know about the distribution of poverty across the region, and the following discussion concentrates on this Model (Table A7), but similar trends are shown in the other Models (Table A6 and Table A8).

Measures of green-ness (specifically the EVI mean for Model 2) and precipitation (igpp51a0) peak at intermediate values of WI, whilst maximum actual evapotranspiration (iget41mx) and the annual amplitude of this variable (iget41a1) both progressively decrease with increases in WI (Table A7). These collectively suggested that the lowest levels of WI are associated with dry areas, intermediate WI levels are associated with moister, greener areas whilst the highest WI levels suggest impacts of human population pressure on the landscape with lower EVI and rainfall values, but with no associated increase in actual evapotranspiration or its variation throughout the year. EVI and rainfall are very strongly correlated ($r^2 = 0.91$ across the 10 WI categories), and it is important to try to understand in which direction this correlation works; does rainfall determine EVI, or EVI determine rainfall? In habitats unaffected by humans, the former must apply, but it is possible that in human-dominated landscapes the latter applies.

Whilst overall model accuracy as determined by the kappa values is relatively low (kappa was 0.344, 0.207 and 0.348 for the three models respectively), the kappa statistic is a very severe judge; classification other than in the correct category is severely penalised regardless of whether or not the miss-classification represented a near-miss or a far-miss. The fourth and fifth columns of the model accuracy figures (lowermost tables) in Tables A6, A7 and A8 show the accuracy figures obtained by allowing errors of plus or minus one or two categories respectively. That is to say, an observation in Category 5 is considered correctly predicted if it is assigned by the model to any category in the range of Categories 4 to 6 (+/- 1 category) or 3 to 7 (+/- 2 categories). There is a considerable increase in model accuracy allowing errors of only +/- 1 category, suggesting that the model fit for all three models is much better than the kappa statistic indicates.

Figure 3. Modelled Wealth Index by household clusters.

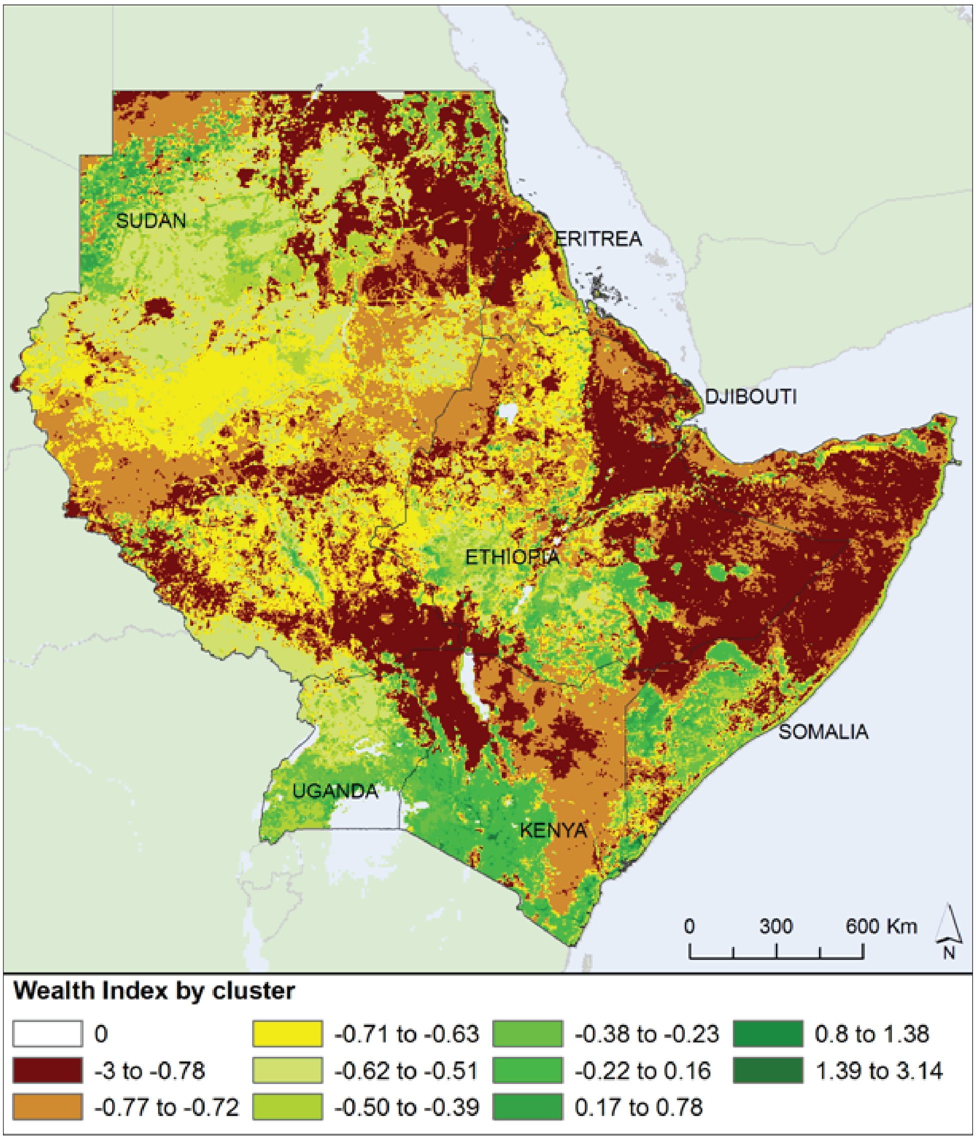


Figure 4. Modelled Wealth Index by individual households.

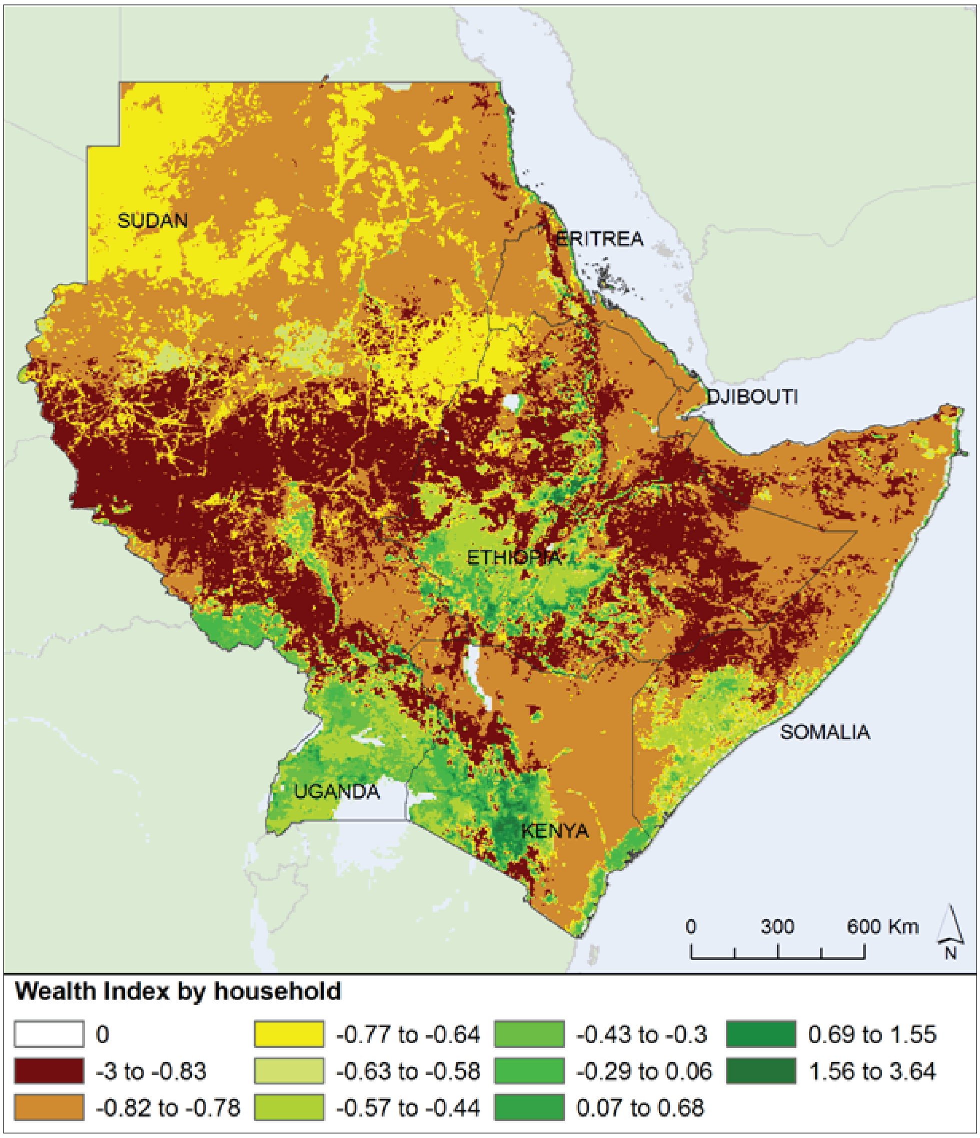
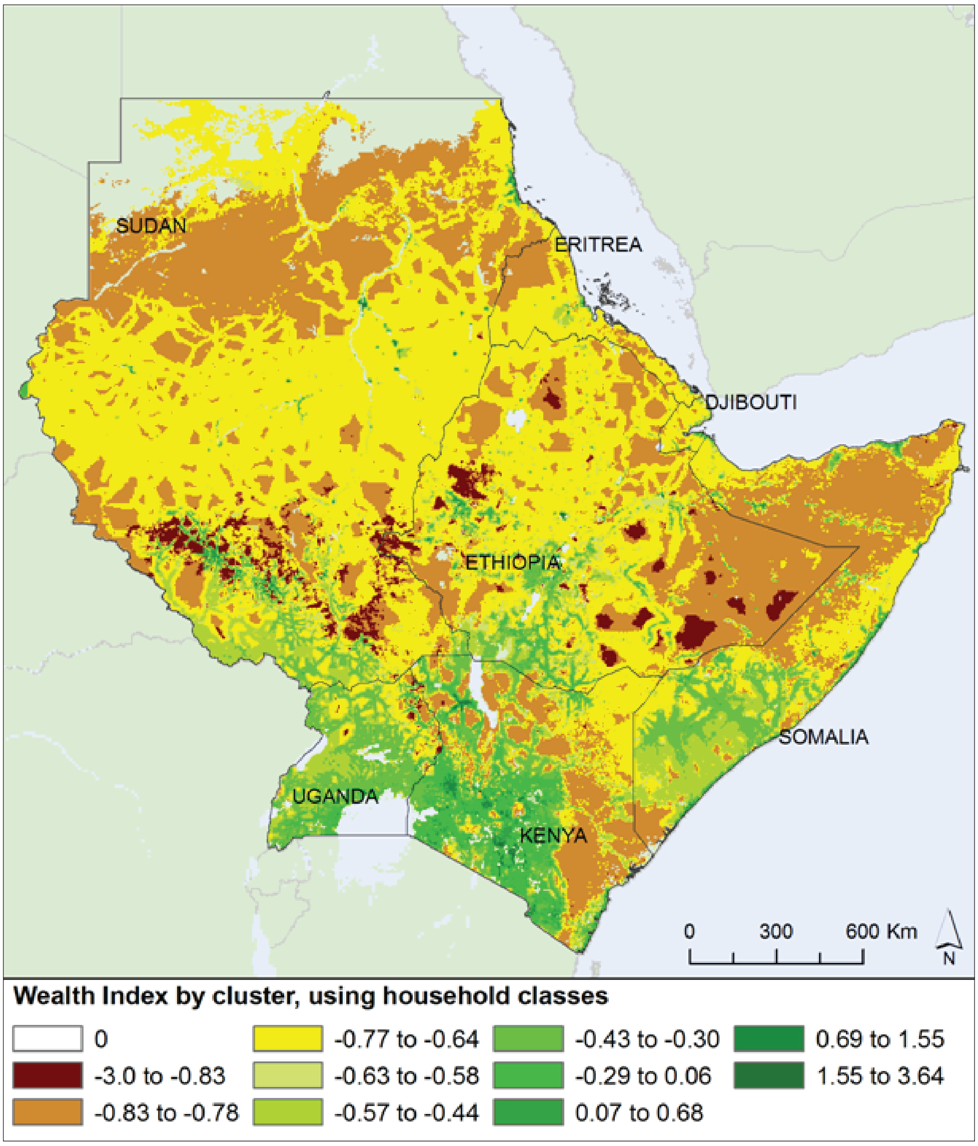


Figure 5. Modelled Wealth Index by household clusters, using the same category boundaries as in Figure 4.



Conclusions

This study confirms the utility of the environmental approach to welfare or poverty mapping, and over a much larger region than heretofore demonstrated (i.e. Uganda). Furthermore the approach appears valid for a region in which there is a much wider range of eco-climatic conditions showing a less obvious trend over the region (as was the case in Uganda, where the dominant eco-climatic trend, and resulting poverty metric, ran from the South-West to the North-East).

The study also shows that it is possible to use the DHS WI as a regional poverty indicator, provided that it is reconstructed from a set of common indicators from the individual DHS surveys.

Obviously there are still some issues to be discussed and steps to undertake. First of all, it is important to test the present models against new field data, or new experts' perspectives of the region. Feedback would tell us both where the current maps are 'right' and where they are 'wrong'.

Whilst the present analysis is region-wide, it is of interest to see if models of the Regional WI at the individual country level are as accurate, or more accurate, than the region-wide model. This will only be possible for the four countries that contributed data to the present exercise; there are no data for the other countries to run the model for them.

As was the case in Uganda, it is important to investigate the scale-dependent accuracy of the current predictions. It is expected that accuracy will increase as the household and satellite data are aggregated into larger geographic units. It is important that this trade-off between accuracy and spatial resolution is resolved at a sufficiently fine spatial unit for the approach to be considered useful by planners, agencies, non-government organisations (NGOs) etc. that are concerned with welfare improvements and poverty alleviation. To know that half a country is poor is of no use if you do not know precisely where the poorest people are located.

High WI values seem to be associated with high human population densities. The reasons for this are difficult to determine with only the data we have at present. It could be because humans acting as individual, free agents congregate in particularly productive areas of the landscape (and consequently enjoy high WI levels), or because aggregations of humans, no matter where they occur, or for what reason, generate sufficient trade and exchange among them so that they collectively enjoy high WI levels regardless of environmental conditions. It appears that these high WI levels are associated with particular sets of environmental conditions, but not the obvious ones that might be expected (e.g. high EVI levels indicating greater photosynthetic activity of all types). Instead the highest WI levels are associated with declining EVI values. Are these lower levels of EVI a cause or a consequence of the high WI values? Has human population pressure in the highest WI areas actually reduced the EVI values through greater or lesser destruction of the natural habitat? Or is it simply that agricultural areas have lower mean EVI values than uncultivated areas (the natural cropping cycle leaves the earth bare for a few months of the year), and the reduction in EVI simply reflects a greater percentage of the ground being brought into the cultivation cycle? These questions raise several key issues about environmental sustainability in areas of highest WI values. We need to know whether high WI values are being achieved at the cost of long-term sustainability.

This paper marks the coming of age of the environmental approach to poverty mapping in Africa. Environmental data may be used to describe welfare across the climatically and sociologically diverse region of the Horn of Africa. One other question remains about this approach. If targeted intervention succeeds in lifting people out of poverty, what changes might we expect to see in the descriptor variables including the environmental signals derived from satellites? For example, cattle densities in Model 3 are at their lowest values in the two highest WI classes. The question is whether these numbers might change as other people currently in lower WI categories (and with more cattle) enter these highest welfare categories. Similarly, this approach raises the question of whether primary production, as indicated by the EVI, will decrease as people currently in the intermediate WI classes (and with high EVI values) become richer.

References

Balk, D. & Yetman, G. 2004. *The global distribution of population: Evaluating the gains in resolution refinement.* Working Paper, CIESIN, Columbia University. Palisades, NY. Available at http://sedac.ciesin.columbia.edu/gpw/docs/gpw3_documentation_final.pdf

Balk, D., Pozzi, F., Yetman, G., Deichmann, U. & Nelson, A. 2004. *The distribution of people and the dimension of place: Methodologies to improve the global estimation of urban extents.* Working Paper, CIESIN, Columbia University. Palisades, NY. Available at http://sedac.ciesin.columbia.edu/gpw/docs/UR_paper_webdraft1.pdf

Center for International Earth Science Information Network (CIESIN), Columbia University, International Food Policy Research Institute (IFPRI), The World Bank & Centro Internacional de Agricultura Tropical (CIAT). 2004. Global Rural-Urban Mapping Project (GRUMP), Beta Version. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw

Center for International Earth Science Information Network (CIESIN), Columbia University & Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World Version 3 (GPWv3). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw

Deaton, A. & Zaidi, S. 2002. *Guidelines for constructing consumption aggregates for welfare analysis.* Living Standards Measurement Study Working Paper No. 135. Washington DC: The World Bank.

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

Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R. & Davis, C.W. 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6): 1373-1379.

FAO. 2006. *Poverty Mapping in Uganda: An analysis using remotely sensed and other environmental data*, by D.J. Rogers, T. Emwanu & T.P. Robinson. Pro-Poor Livestock Policy Initiative (PPLPI) Working Paper No. 36. Rome.

FAO. 2007. *Gridded Livestock of the World,* by G.R.W. Wint & T.P. Robinson. Rome. 145 pp. Available at www.fao.org/ag/AGAinfo/resources/en/glw/default.html

FAO. 2008. *Poverty and welfare measures in the Horn of Africa*, by F. Pozzi & T.P. Robinson. IGAD LPI Working Paper No. 8. Rome.

Field, A.P. 2005. *Discovering statistics using SPSS* (2nd. Edition). London: Sage.

Filmer, D. & Pritchett, L.H. 2001. Estimating wealth effects without expenditure data – or tears: An application to educational enrolments in states of India. *Demography*, 38: 115-132.

Hentschel, I., 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.

Montana, L. & Spencer, J. 2004. *Incorporating geographic information into MEASURE Surveys: A field guide to GPS data collection.* Carolina Population Center Report, 25 February 2004.

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

Rogers, D.J. and Williams, B.G. 1994. Tsetse distribution in Africa: Seeing the wood and the trees. In: Edwards, P.J., May, R., Webb, N.R., eds (1994) *Large-scale ecology and conservation biology*, pp 247–271. Oxford: Blackwell Scientific.

Rutstein, S.O. & Johnson, K. 2004. *The DHS Wealth Index*. DHS Comparative Reports No. 6. Calverton, Maryland: ORC Macro.

Rutstein, S.O. & Rojas, G. 2003. *Guide to DHS Statistics.* Demographic and Health Surveys. Calverton, Maryland: ORC Macro.

Sanderson, E.W., Jaiteh, M. Levy, M.A., Redford, K.H., Wannebo, A.V. & Woolmer, G. 2002. The Human Footprint and The Last of the Wild. *BioScience*, 52, n.10 (October 2002): 891-904.

Scharlemann, J.P.W., Benz, D., Hay, S.I., Purse, B.V., Tatem, A.J., Wint, G.R.W. & Rogers, D.J. 2008. Global data for ecology and epidemiology: A novel algorithm for temporal Fourier processing MODIS data. *PLoS ONE*, 3(1): e1408. doi:10.1371/journal.pone.0001408.

Wildlife Conservation Society (WCS) and Center for International Earth Science Information Network (CIESIN). 2002. Last of the Wild Data Version 1 (LWP-1): Global Human Footprint data set (HF).

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

World Bank. 2003. *Measuring living standards: Household consumption and wealth indices, quantitative techniques for health equity analysis.* Technical Note # 4. Washington DC.

Annex 1: Tables

In this section we report the correlations between the different wealth indices, in particular the estimated Regional WI, and a set of wealth indices derived from the original DHS data (Tables A2 to A5). The critical correlation between the original country-specific WI using all the variables available for that country, and the Regional WI using just the variables common to all country datasets is highlighted in red. The correlation between the country-specific WI calculated using only the common regional variables and the same Regional WI is highlighted in blue.

Tables A6, A7 and A8 show the results of the model (at the cluster and household level), through the list of the predictor variables and their means, and the accuracy results.

Table A 1. Definitions of the wealth indices calculated for this project.

Name	Definition
WI_COMMC	Country-specific WI calculated with common variables
WI_ALLC	Country-specific WI calculated with all available variables
WI_REG	Regional WI (with all common variables)
WI_REGIM	Regional WI without wc_bush and fl_dirt
WI_REG2	Regional WI without w_pipe_r and wc_pit
WLTHINF	Original WI

Table A 2. Correlations between the different wealth indices for Eritrea. All correlation values are significant at p<0.001 or better.

Name	ERITREA 2002	WI_COMMC	WI_ALLC	WI_REG	WI_REGIM	WI_REG2	WLTHINF
WI_COMMC	Pearson Correlation	1	0.994	0.995	0.983	0.994	0.973
	N	9 389	7 194	9 389	9 389	9 389	9 389
WI_ALLC	Pearson Correlation	0.994	1	0.99	0.983	0.99	0.98
	N	7 194	7 194	7 194	7 194	7 194	7 194
WI_REG	Pearson Correlation	0.995	0.99	1	0.99	1	0.969
	N	9 389	7 194	9 389	9 389	9 389	9 389
WI_REGIM	Pearson Correlation	0.983	0.983	0.99	1	0.991	0.959
	N	9 389	7 194	9 389	9 389	9 389	9 389
WI_REG2	Pearson Correlation	0.994	0.99	1	0.991	1	0.968
	N	9 389	7 194	9 389	9 389	9 389	9 389
WLTHINF	Pearson Correlation	0.973	0.98	0.969	0.959	0.968	1
	N	9 389	7 194	9 389	9 389	9 389	9 389

Table A 3. Correlations between the different wealth indices for Ethiopia. All correlation values are significant at p<0.001 or better.

Name	ETHIOPIA 2005	WI_COMMC	WI_ALLC	WI_REG	WI_REGIM	WI_REG2	WLTHINF
WI_COMMC	Pearson Correlation	1	0.987	0.996	0.973	0.992	0.962
	N	13 721	13 721	13 721	13 721	13 721	13 721
WI_ALLC	Pearson Correlation	0.987	1	0.985	0.966	0.981	0.97
	N	13 721	13 721	13 721	13 721	13 721	13 721
WI_REG	Pearson Correlation	0.996	0.985	1	0.987	0.999	0.965
	N	13 721	13 721	13 721	13 721	13 721	13 721
WI_REGIM	Pearson Correlation	0.973	0.966	0.987	1	0.991	0.954
	N	13 721	13 721	13 721	13 721	13 721	13 721
WI_REG2	Pearson Correlation	0.992	0.981	0.999	0.991	1	0.964
	N	13 721	13 721	13 721	13 721	13 721	13 721
WLTHINF	Pearson Correlation	0.962	0.97	0.965	0.954	0.964	1
	N	13 721	13 721	13 721	13 721	13 721	13 721

Table A 4. Correlations between the different wealth indices for Kenya. All correlation values are significant at p<0.001 or better.

Name	KENYA 2003	WI_COMMC	WI_ALLC	WI_REG	WI_REGIM	WI_REG2	WLTHINF
WI_COMMC	Pearson Correlation	1	0.981	0.988	0.987	0.993	0.928
	N	8 561	8 561	8 561	8 561	8 561	8 561
WI_ALLC	Pearson Correlation	0.981	1	0.971	0.977	0.975	0.923
	N	8 561	8 561	8 561	8 561	8 561	8 561
WI_REG	Pearson Correlation	0.988	0.971	1	0.982	0.999	0.916
	N	8 561	8 561	8 561	8 561	8 561	8 561
WI_REGIM	Pearson Correlation	0.987	0.977	0.982	1	0.986	0.898
	N	8 561	8 561	8 561	8 561	8 561	8 561
WI_REG2	Pearson Correlation	0.993	0.975	0.999	0.986	1	0.917
	N	8 561	8 561	8 561	8 561	8 561	8 561
WLTHINF	Pearson Correlation	0.928	0.923	0.916	0.898	0.917	1
	N	8 561	8 561	8 561	8 561	8 561	8 561

Table A 5. Correlations between the different wealth indices for Uganda. All correlation values are significant at p<0.001 or better.

Name	UGANDA 2001	WI_COMMC	WI_ALLC	WI_REG	WI_REGIM	WI_REG2	WLTHINF
WI_COMMC	Pearson Correlation	1	0.978	0.984	0.981	0.987	0.953
	N	7 885	7 885	7 885	7 885	7 885	7 885
WI_ALLC	Pearson Correlation	0.978	1	0.972	0.96	0.973	0.963
	N	7 885	7 885	7 885	7 885	7 885	7 885
WI_REG	Pearson Correlation	0.984	0.972	1	0.974	0.999	0.952
	N	7 885	7 885	7 885	7 885	7 885	7 885
WI_REGIM	Pearson Correlation	0.981	0.96	0.974	1	0.98	0.932
	N	7 885	7 885	7 885	7 885	7 885	7 885
WI_REG2	Pearson Correlation	0.987	0.973	0.999	0.98	1	0.951
	N	7 885	7 885	7 885	7 885	7 885	7 885
WLTHINF	Pearson Correlation	0.953	0.963	0.952	0.932	0.951	1
	N	7 885	7 885	7 885	7 885	7 885	7 885

Table A 6. Modelled WI by household cluster (Model 1) details. Predictor variables and their mean values are shown in the upper two tables. Model accuracy by category is shown in the lower table.

WI_REG	hum foot	Ig1515a0	igpp51p2	ig1507a0	iget41mx	igep42mx	ig1514p1	ig0535mn	iget41a3	ig1508p3	n (Sample)
Cat. 1	23.38	0.21	2.6	35.59	269.99	467.87	7.57	7.79	31.43	1.67	151
Cat. 2	26.1	0.22	2.53	34.63	274.3	451.77	7.43	11.05	32	1.8	159
Cat. 3	27.51	0.25	2.6	33.22	267.96	447.36	7.63	22.4	28.32	1.63	149
Cat. 4	28.82	0.27	2.74	32.57	272.95	440.37	7.62	35.34	27.47	1.9	148
Cat. 5	31.94	0.35	3.16	30.84	259.26	434.01	6.41	69.66	24.18	2.05	155
Cat. 6	33.33	0.36	3.4	30.48	253.25	455.67	6.44	79.07	28.75	1.92	151
Cat. 7	33.93	0.36	3.69	29.74	247.46	469.81	5.5	67.97	29.26	2.07	151
Cat. 8	43.67	0.31	3.55	30.66	238.93	461.84	5.79	57.13	28.28	1.8	152
Cat. 9	51.83	0.25	3.17	31.65	242.95	444.36	6.31	29.24	27.52	1.79	148
Cat. 10	63.33	0.2	3.15	31.01	220.99	441.58	5.61	4.67	31.07	1.55	150

Name	Variable Name
humfoot	Human Footprint
ig1515a0	EVI mean
igpp51p2	EARSNL Precipitation phase2
ig1507a0	Day LST mean
iget41mx	EARSNL Actual Evapotranspiration maximum
igep42mx	EARSNL Potential Evapotranspiration maximum
ig1514p1	NDVI phase1
ig0535mn	Evapotranspiration minimum
iget41a3	EARSNL Actual Evapotranspiration amp3
ig1508p3	Night LST phase3

Category	Description	% Correct	% Correct (+/-1 cat.)	% Correct (+/-2 cat.)	% Producer's Accuracy	% Consumer's Accuracy
Cat. 1	-3.0 to -0.8	60.3	72.2	88.7	60.3	45.7
Cat. 2	-0.8 to -0.7	34	78.6	91.8	34	32.9
Cat. 3	-0.7 to -0.6	45.6	69.8	83.9	45.6	33.7
Cat. 4	-0.6 to -0.5	29.1	49.3	75.7	29.1	30.7
Cat. 5	-0.5 to -0.4	27.7	55.5	78.1	27.7	50.6
Cat. 6	-0.4 to -0.2	35.1	64.2	76.2	35.1	35.6
Cat. 7	-0.2 to 0.2	49.7	67.5	77.5	49.7	35.5
Cat. 8	0.2 to 0.8	20.4	59.2	80.3	20.4	44.9
Cat. 9	0.8 to 1.4	31.1	69.6	81.8	31.1	46.5
Cat. 10	1.4 to 3.1	77.3	85.3	88	77.3	59.2

Table A 7. Modelled WI by individual household (Model 2) details. Predictor variables and their mean values are shown in the upper two tables. Model accuracy by category is shown in the lower table.

WI_REG	hum foot	ig1515a0	ig0535p2	igpp51a0	ig1508a2	iget41a1	igep42mx	iget41mx	ig1507p3	iget41a2	n (Sample)
Cat. 1	26.25	0.25	2.56	223.83	1.09	63.72	453.64	267.52	1.71	50.66	4713
Cat. 2	26.17	0.21	2.37	191.86	1.33	68.69	457.23	270.29	1.63	54.19	4463
Cat. 3	28.66	0.26	2.59	232.18	1.2	66.34	452.48	268.62	1.74	52.5	2876
Cat. 4	28.65	0.25	2.53	216.76	1.17	62.54	443.74	263.56	1.62	53.58	2797
Cat. 5	31.9	0.36	3.3	307.37	0.73	46.09	444.7	259.59	1.87	41.81	3957
Cat. 6	33.4	0.35	3.14	294.8	0.83	48.18	453.85	259.76	1.8	43.28	3462
Cat. 7	35.8	0.34	3.06	288	0.79	45.59	452.99	255.34	1.78	43.56	3761
Cat. 8	42.3	0.31	2.95	274.16	0.76	39.2	451.31	247.2	1.65	45.99	3675
Cat. 9	51.09	0.27	2.81	261.57	0.74	34.29	448.85	235.96	1.56	47.89	3687
Cat. 10	59.41	0.23	2.82	237.23	0.74	34.99	448.88	229.82	1.45	53.43	3587

Name	Variable Name
humfoot	Human Footprint
ig1515a0	EVI mean
ig0535p2	ETR phase2
igpp51a0	EARSNL Precipitation mean
ig1508a2	Night LST amp2
iget41a1	EARSNL Actual Evapotranspiration amp1
igep42mx	EARSNL Potential Evapotranspiration maximum
iget41mx	EARSNL Actual Evapotranspiration maximum
ig1507p3	Day LST phase3
iget41a2	EARSNL Actual Evapotranspiration amp2

Category	Description	% Correct	% Correct (+/-1 cat.)	% Correct (+/-2 cat.)	% Producer's Accuracy	% Consumer's Accuracy
Cat. 1	-3.0 to -0.8	47.3	70.9	72.8	47.3	30.7
Cat. 2	-0.8 to -0.8	44	76.9	81.6	44	32.3
Cat. 3	-0.8 to -0.6	4.6	36.1	86.3	4.6	19.4
Cat. 4	-0.6 to -0.6	8.9	31.5	66.2	8.9	21.2
Cat. 5	-0.6 to -0.4	39.5	54.6	60.1	39.5	22.9
Cat. 6	-0.4 to -0.3	17.3	53.2	65	17.3	20.6
Cat. 7	-0.3 to 0.1	8.1	32.9	65.3	8.1	24.2
Cat. 8	0.1 to 0.7	16	35.6	65.9	16	22.7
Cat. 9	0.7 to 1.5	16.7	70.9	75	16.7	28.1
Cat. 10	1.6 to 3.6	71.7	81.3	87.6	71.7	42.9

Table A 8. Modelled WI by household cluster, but using the category boundaries from Model 2 (Model 3) details. Predictor variables and their mean values are shown in the upper two tables. Model accuracy by category is shown in the lower table.

WI_REG	gmppop	dallrd	ig1514p2	ig1507p1	mktaces	ig1507p2	iget41a0	igpp51p2	Cattle	GLC2K	n (Sample)
Cat. 1	47.81	7.78	3.25	2.11	8.24	3.3	158.37	2.81	67.66	12.63	41
Cat. 2	53.72	7.78	2.88	3.13	8.9	3.55	135.02	2.52	37.51	13.88	109
Cat. 3	81.89	4.5	2.91	2.78	6.73	3.38	140.26	2.52	48.43	13.06	294
Cat. 4	110.07	4.7	2.96	2.77	6.32	3.24	147.78	2.62	62.81	13.09	69
Cat. 5	177.43	2.89	3.78	2.69	5.34	2.78	161.73	3	55.21	11.99	187
Cat. 6	243.91	1.52	4.09	2.85	3.53	2.57	174.64	3.31	61.2	12.76	130
Cat. 7	343.5	1.2	4.37	3.77	2.82	2.46	158.11	3.62	68.23	12.26	178
Cat. 8	1 155.95	0.5	4.31	4.16	2.35	2.66	157.04	3.57	46.49	12.97	144
Cat. 9	3 848.92	0.47	3.64	3.46	0.94	2.94	154.72	3.21	29.62	16.39	187
Cat. 10	7 062.34	0.58	3.79	5.1	0.73	3.05	138.27	3.18	24.12	20.27	127

Name	Variable Name
gmppop	GRUMP Population Density
dallrd	Distance to all roads
ig1514p2	NDVI phase2
ig1507p1	Day LST phase1
mktaces	Access to markets
ig1507p2	Day LST phase2
iget41a0	EARSNL Actual Evapotranspiration mean
igpp51p2	EARSNL Potential Evapotranspiration phase2
Cattle	Cattle Density
GLC2K	Global Land Cover 2000

Category	Description	% Correct	% Correct (+/-1 cat.)	% Correct (+/-2 cat.)	% Producer's Accuracy	% Consumer's Accuracy
Cat. 1	-3.0 to -0.8	14.6	36.6	95.1	14.6	42.9
Cat. 2	-0.8 to -0.8	31.2	92.7	92.7	31.2	41
Cat. 3	-0.8 to -0.6	74.8	86.1	92.2	74.8	43.4
Cat. 4	-0.6 to -0.6	2.9	75.4	94.2	2.9	12.5
Cat. 5	-0.6 to -0.4	21.4	40.6	94.1	21.4	36.7
Cat. 6	-0.4 to -0.3	40	76.2	80.8	40	33.1
Cat. 7	-0.3 to 0.1	44.4	69.1	81.5	44.4	38
Cat. 8	0.1 to 0.7	23.6	63.9	85.4	23.6	37.4
Cat. 9	0.7 to 1.6	48.1	78.6	90.9	48.1	61.6
Cat. 10	1.6 to 3.6	71.7	81.3	87.6	71.7	42.9

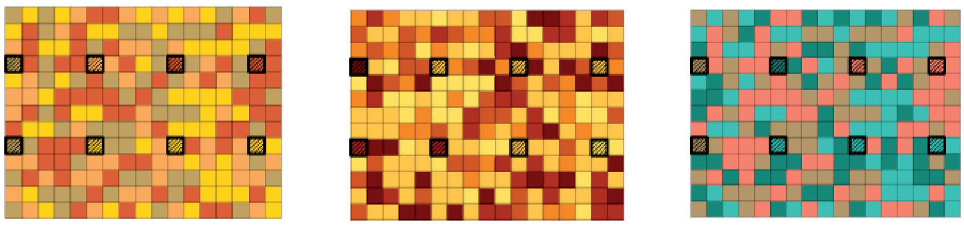
Annex 2: Livestock modeling

The predictor layers used in the WI mapping exercise for cattle, camel, chicken, pig, goat and sheep densities and cropping percentage were themselves derived by modelling a set of point or administrative level data for these variables. This Annex describes how this was done.

The underlying process of livestock and crop distribution modelling is covered extensively in FAO (2007). Once the available agricultural statistics have been collected, standardised, enhanced with supplementary data and adjusted for the extent of land deemed suitable for livestock production, the resulting data provide a sound base for statistical distribution modelling. The model then relies on the use of raster images to store both observed (or training) data (i.e. livestock densities) and all the predictor variables. Statistical relationships are established between observed and predictor variables using values extracted for a series of regularly spaced sample points, as illustrated in Figure A1. The resulting equations are then applied to all the pixels in the predictor images, to produce a predicted distribution map.

Figure A 1. Schematic livestock and crop distribution modeling.

Observed density - A x (Predictor 1) + B x (Predictor2)



Step 1: Convert all data maps to images with same pixel size (resolution)

Step 2: Extract values for observed values of livestock density, and for each predictor variable at fixed sample points (hatched squares)

*Step 3: Calculate a regression equation of the form: Observed density = Constant + A * (Predictor 1) + B * (Predictor 2) + …*

The technique can therefore be used to predict livestock or crop distributions in areas for which no livestock data are available, i.e. filling in gaps. Moreover, because predicted distributions are produced at the resolution of the raster imagery, the models generate heterogeneous distributions within polygons that have only a single observed value, thus disaggregating the original data. For limited datasets therefore, the method has the major advantage of both filling in gaps and refining the level of detail that can mapped.

A wide variety of predictor variables is used in the modelling process, embracing environmental, demographic, climatic, agricultural, topographic, and infrastructural factors. The majority of environmental and climatic parameters are derived from either public domain global datasets (elevation, land use and land cover, human population), from GIS processing (distance to features such as roads and towns) or from the MODIS satellite imagery referred to in the main report.

It is by no means certain that relationships between target and predictor variable are linear, and it is therefore advisable to test non linear relationships. This is achieved by numerically transforming the variable values prior to statistical analysis. Models were thus assessed with dependent and independent variables in both their un-transformed state, and with the natural logarithmic transformation applied.

The predictors of animal density are also unlikely to be consistent from region to region, and the modelling process should therefore be run at several different spatial scales to provide a range of predictive relationships appropriate to specific areas. As well as administrative level analyses, an ecological stratification was used on the assumption that the factors determining animal distributions are likely to be similar in areas with comparable ecological characteristics, thereby allowing a) more robust statistical relationships to be established between training data and predictor variables and b) more realistic predictions of livestock densities in other parts of the same analysis zone for which data are not available.

The modelled outputs for cattle, camel, chicken, pig, goat and sheep densities, and for cropping percentage, from the Gridded Livestock of the World (GLW), are illustrated in Figure A2 and Figure A3, respectively.

Figure A 2. Livestock densities: a) cattle, b) camel, c) chicken, d) pig, e) goat and f) sheep.

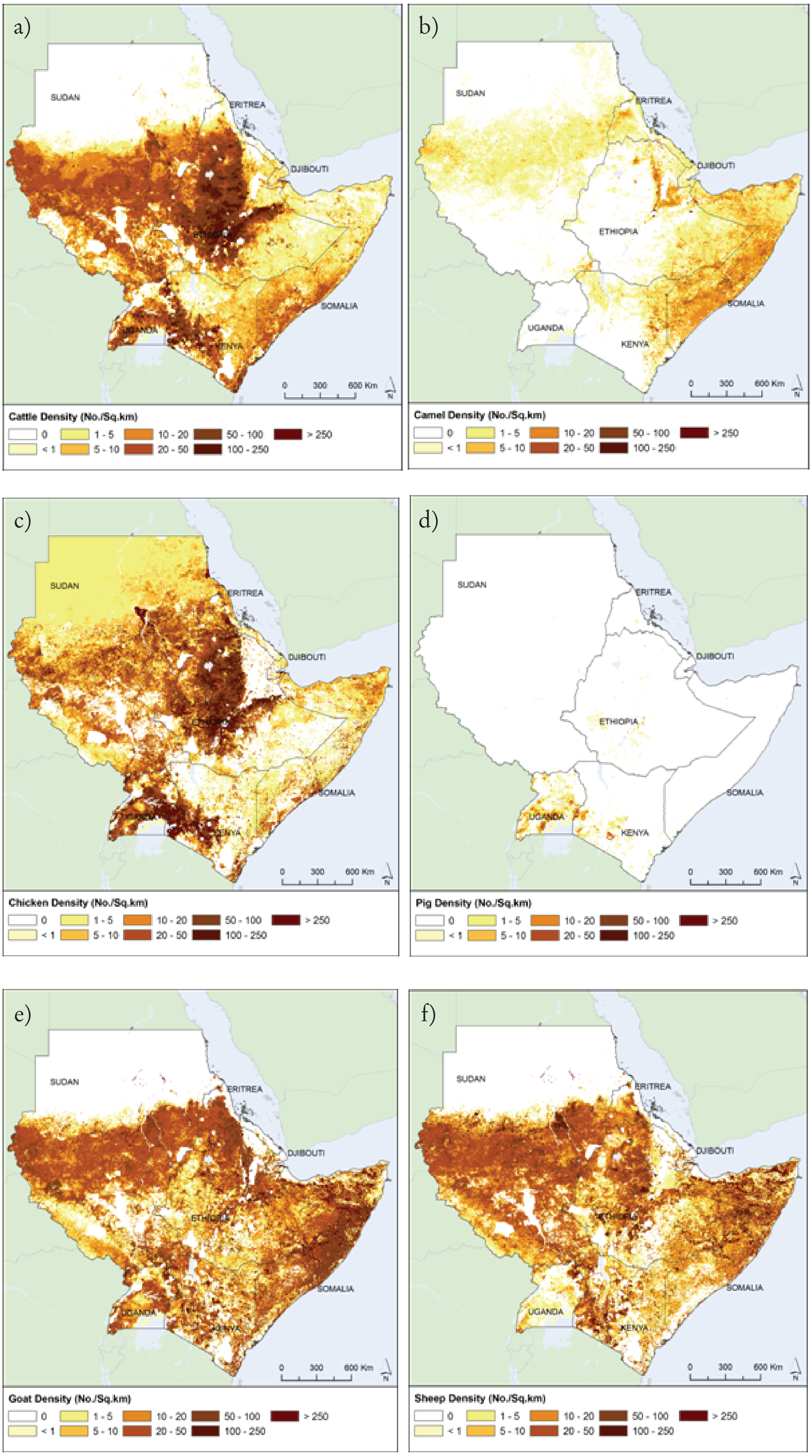


Figure A 3. Cropping percentage.