

Published in final edited form as:

Trop Med Int Health. 2006 August ; 11(8): 1318–1326. doi:10.1111/j.1365-3156.2006.01681.x.

Mapping the environmental coverage of the INDEPTH demographic surveillance system network in rural Africa

Andrew J. Tatem¹, Robert W. Snow^{2,3}, and Simon I. Hay^{1,2}

¹Department of Zoology, Spatial Ecology and Epidemiology Research Group, University of Oxford, Oxford, UK

²Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, Nairobi, Kenya

³Centre for Tropical Medicine, John Radcliffe Hospital, University of Oxford, UK

Summary

OBJECTIVES—The INDEPTH DSS network was founded in 1998 to provide an international network of field sites for continuous demographic evaluation of populations and their health. Results from the network have been used to derive estimates of mortality, morbidity and health equity. Spatial extrapolation and logical summaries of these findings are dependent on the network covering a representative sample of the environments in a region and their interrelationships being known. Here, we investigate how comprehensive is the coverage of the network of rural DSS sites in Africa in terms of the range of ecological zones found across the continent.

METHODS—We used satellite imagery to define an environmental signature for each INDEPTH DSS site, and then calculate Euclidean distances from these signatures to the environmental signatures of every image pixel across Africa. These distances were then mapped and a gridded population surface used to mask uninhabited areas to illustrate the extent of the environmental coverage of the INDEPTH network. Environmental similarities between DSS sites were also calculated, hierarchically clustered and visualized as a dendrogram to examine between site relationships. Finally, an ecozonation of Africa was used to analyse the per-ecozone environmental similarity of the INDEPTH DSS network.

RESULTS AND CONCLUSIONS—The current INDEPTH DSS network in Africa spans all the major environmental zones, but within these zones the environmental coverage of the network varies. These variations were mapped by ecozone. These maps provide valuable information in determining the confidence with which relationships derived from rural INDEPTH DSS sites can be extended to other areas. The results also indicate suites of sites that form environmentally cohesive groups and from which data can be logically summarized. Finally, the results highlight areas where the location of new INDEPTH DSS sites would increase significantly the environmental coverage of the network.

Keywords

demographic surveillance; INDEPTH network; satellite imagery; environmental distance; ecozonation; dendrogram

Introduction

Since the early 1960s, an increasing number of research groups in developing countries have established longitudinal data collection systems to circumvent unreliable national civil registration systems for estimates of age- and cause-specific mortality and their determinants. These data collection systems are known as demographic surveillance systems (DSS) and are defined as ‘the longitudinal follow-up of well-defined entities or primary subjects (individuals, households and residential units) and all related demographic and health outcomes within a clearly circumscribed geographical area’ (INDEPTH 2002). A DSS involves an initial population census, followed by regular registrations of births, deaths, migrations and other data (such as morbidity episodes, pregnancies, marriages and economic activities), usually at 3–4 month intervals (INDEPTH 2005). In 1998, many research DSS sites united to found the International Network of field sites with continuous Demographic Evaluation of Populations and Their Health (INDEPTH) (Ngom *et al.* 2001). This INDEPTH DSS network includes 25 sites in sub-Saharan Africa (SSA) (http://www.indepth-network.net/dss_site_profiles/dss_sites.htm).

One of the principal purposes of the INDEPTH network is to provide representative data on a range of demographic and health indicators in low income regions of the world, where such information is scarce. Moreover, INDEPTH aims to pool these site data to provide information on the wide-area demographic patterns and their determinants (INDEPTH 2005). Morbidity, mortality and health equity measures in Africa, however, are influenced strongly by the burden of infectious diseases (Murray & Lopez 1997). Many of these diseases are sensitive to environmental conditions, particularly those with intermediate hosts in their life cycle (Hay 2000). Subnational economies and poverty, determinants of disease outcomes, are also governed by ecology and climate (Gallup & Sachs 2001). To enable confident comparison between the DSS sites and the effective extension of findings to other areas, the environmental characteristics of the sites should therefore be similar. Information on the environmental grouping of the 25 DSS sites in SSA should be used in the optimal summaries of health characteristics of populations in this region.

In this article, we examine how far, and with what level of confidence, data generated from African DSS sites can be extended, what sites can be grouped and also where additions to the INDEPTH DSS network would maximize environmental coverage. Analysis is conducted within zones of ecological and climatic similarity (ecozones; FAO 2000; Olson *et al.* 2001; Strahler & Strahler 2002; Schultz 2005) to maintain data consistency in terms of environmental conditions and, therefore, provide both focussed and relevant information on network coverage and site intercomparability. As drivers of urban health are arguably less closely coupled to the environment (Tatem & Hay 2004; Hay *et al.* 2005a), we focus on the network of rural DSS sites. We also investigate only Africa ($n = 25$), as the site coverage in Asia ($n = 10$) and South America ($n = 1$) is relatively sparse.

Methods

Data

Of the 25 DSS sites within the African INDEPTH network in 2005 (details of each can be found at http://www.indepth-network.net/dss_site_profiles/dss_sites.htm), 21 are located partly or wholly in rural areas. These 21 DSS site locations were digitized from maps provided in INDEPTH (2002), except where administrative boundary data corresponded exactly and could thus be imported directly from other sources (http://www3.who.int/whosis/gis/salb/salb_home.htm; <http://www.fao.org/geonetwork>).

Environmental variables were derived from the 8×8 km Global Area Coverage Advanced Very High Resolution Radiometer (AVHRR) imagery from the National Oceanographic and Atmospheric Administration's series of polar-orbiting Television Infrared Observation Satellites (Hay *et al.* 2006). The available decadal imagery sequence from 1982 to 1999 was maximum value composited into monthly files and then subjected to Temporal Fourier Analysis (Rogers 2000; Rogers *et al.* 2002). Of the outputs, the minimum, maximum and mean middle infrared radiation (MIR), land surface temperature (LST) and normalised difference vegetation index (NDVI) images were used. Together, these three variables provide a relatively complete picture of environmental conditions in an area (Hay & Lennon 1999).

Middle infrared radiation is recorded by channel 3 of the AVHRR and is correlated with the water content, surface temperature and structure of vegetation canopies (Boyd & Curran 1998). LST images were calculated from AVHRR channels 4 and 5 using a simple split-window algorithm (Price 1984) and have been shown to exhibit accuracies equivalent to spatial interpolation of meteorological data (Hay & Lennon 1999). The NDVI is widely used within the field of remote sensing, and is specifically a measure of chlorophyll abundance, but is also correlated with soil moisture, rainfall and vegetation biomass, coverage and productivity (Campbell 1996).

The data, findings and relationships derived from sites within the INDEPTH network will obviously not be of practical value in uninhabited regions. Therefore, the Global Rural–Urban Mapping Project population layer (Balk *et al.* 2006) was used to produce a mask to exclude from analysis any areas inhabited by <1 person per km^2 , as it has been shown to be the most reliable population surface for the region (Hay *et al.* 2005b).

Environmental signatures and distances

The area covered by each INDEPTH DSS site was overlaid on the nine satellite image layers, each rescaled to the same data range, and the average value of each satellite variable at the site was extracted. Thus, for each DSS site, minimum, maximum and mean rescaled MIR, LST and NDVI values were obtained. These nine values represented a unique 'environmental signature' for each INDEPTH DSS site.

To examine how similar the environmental signature of each INDEPTH DSS site was to the environments throughout the rest of Africa, 'environmental distance' measures were calculated. These distances provided a continuous measure of environmental similarity, with low values representing similar environments and high values more dissimilar environments. The lack of sufficient variance in the majority of INDEPTH DSS site signatures (as the areas covered by most sites were insufficient to include large numbers of 8×8 km AVHRR imagery pixels) dictated that only 'Euclidean' distances could be used as a measure of environmental similarities. Euclidean distance is defined as the shortest straight line distance between two points, in this case, the distance between the environmental signature centroids in nine-dimensional environmental space (ERDAS 2003), as defined by the Fourier-processed satellite images (Hay *et al.* 2006). Each dimension is defined by the full range of values taken by each rescaled satellite-derived environmental variable, and within this space a single point (or centroid) represents, for example, those specific values at Kisumu in Kenya (its environmental signature). Elsewhere in this nine-dimensional space, a different point represents the environmental conditions at a pixel located in southern Chad. The Euclidean distance within this space from the Kisumu point to the Chad point represents the environmental distance between the two locations.

Environmental distances from each DSS site environmental signature to every satellite image pixel were calculated and mapped. The resulting image was a continuous raster layer,

in which each pixel value represented the Euclidean distance between the pixel and the mean vector of the DSS site signatures. The pixels with the higher Euclidean distance values were environmentally more different from the signature means. The pixels with low distance values were environmentally more similar, with the lowest value pixels, by definition, found at the DSS sites. Note that these values are rarely zero, as they represent the difference from the average environmental signature of the sites, and almost all sites cover more than one image pixel.

Environmental classification and hierarchical clustering

In extending findings from individual DSS sites to other locations and when wishing to summarize, data consistency in environmental conditions is desirable. For example, results derived from the highly vegetated and humid environment at the Kisumu DSS cannot realistically be extended to or summarized with those from the dry, Sahelian climate of Nouna DSS. A basic ecological zonation was undertaken, therefore, to identify the broad, environmental grouping across Africa (Schultz 2005).

The classification used the same nine rescaled satellite variable images defined earlier. Iterative Self-Organizing Data Analysis Technique (ISODATA) clustering (ERDAS 2003) in multivariate space was used to map 50 environmental classes. These represented areas where, within classes, environmental and climatic regimes were similar, but between classes this similarity was minimized. Separability and dendrogram analysis (ERDAS 2003) was used to examine the between-class similarities and to determine whether and how the 50 classes could be grouped into a lower number of classes, each environmentally distinct from one another. The separability analysis involved the examination of changes in environmental distances through the grouping of environmental classes, while the dendrogram (described below) provided a visual representation of the environmental links between classes to aid the direction and limits of grouping. The analysis focussed on the maximization of Mahalanobis distance (Webb 2003) between clusters and revealed that the 50 classes still encompassed significant between-class similarity and could be most parsimoniously grouped into just four distinct classes. This four-class grouping also matched numerous previous environmental and climatic classifications of Africa, which have found four or five distinct zones on the continent, each with unique levels of vegetation and temperature, precipitation and humidity regimes (FAO 2000; Olson *et al.* 2001; Strahler & Strahler 2002; Schultz 2005). The four classes matched those described elsewhere as class 1 = desert and semidesert, class 2 = thorn savannah and steppes, class 3 = savannah with summer rain, class 4 = tropics with year round rain. Per-class environmental distance images were then calculated, using only the DSS signatures belonging to the class in question. In order to avoid unrealistic discrete class boundaries and an over-reliance on the ecozonation results, multiple class membership was incorporated if a site was located on or within 100 km of class boundaries.

To examine environmental similarities between sites, the environmental distance between each site signature and every other signature was calculated. These were formed into an 'environmental distance dissimilarity matrix'. Hierarchical clustering procedures are the most commonly used method of summarizing such data structures (Webb 2003). The clustering process produces a hierarchical tree, which is a nested set of partitions represented by a dendrogram (tree diagram). Dendrograms are visual representations of the results of hierarchical clustering that are commonly used in the fields of evolution and genetics, and occasionally in an environmental context (Sugihara *et al.* 2003; Rogers & Robinson 2004; Tatem *et al.* 2006a,b). Here, dendrograms were used to provide an environmental representation of the INDEPTH DSS network in Africa by demonstrating visually the environmental relations between each site. The clustering results were translated into dendrograms based on centroid linkage (Webb 2003). Those sites that are adjacent and linked by low branches are environmentally similar, while those linked by long branches are

more different. A simple test using other distance measures where possible [e.g. Mahalanobis, divergence, Jefferies-Matusita (ERDAS 2003; Webb 2003)] showed no obvious changes in the dendrogram architecture from the Euclidean one used here.

Results

Figure 1 shows a three-dimensional perspective map of overall environmental distance from the DSS sites. It demonstrates which areas were relatively most different environmentally from the current INDEPTH network of rural sites in Africa. It is clear that the most hot and arid regions south of the Sahara are least well represented by the network. This includes the northern Sahel region, the horn of Africa and northern Kenya. Elsewhere, eastern Madagascar and western Angola also show relatively large environmental distances from the network.

Figure 2a shows the four environmental classes or ecozones with areas of <1 person per km² masked. Figure 2b summarizes the characteristics of the four environmental classes, in terms of minimum, maximum and mean value of each of the satellite-derived environmental variables: hot and sparsely vegetated (class 1), relatively cool and highly vegetated (class 4) and temperatures and vegetation levels falling between the two (classes 2 and 3). The application of a low-density population mask reduced the class areas, particularly for classes 1 and 2, which generally represent inhospitable environments to humans.

Figure 3 shows the dendrogram produced from hierarchical clustering of the environmental variables at each DSS site and provides a visual representation of the environmental similarity structure and linkages within the current INDEPTH network. The environmental class memberships marked on the dendrogram show that distinct groupings of sites with similar environments appear, some spanning the border between classes.

Figure 4 maps environmental distances per environmental class. These per-class distance maps provide more focussed information on coverage of the network, confidence in the extrapolation of findings from individual INDEPTH network sites and guides as to future DSS site locations that would improve per-class environmental coverage optimally.

Discussion

Network coverage

The existing rural INDEPTH DSS network for Africa cannot be described as either environmentally comprehensive or incomplete using these methods, but estimates of relative levels of coverage can be calculated. Figure 1 is best interpreted as an overall summary environmental coverage for Africa and a useful tool for gauging uncertainty in extrapolations. The large distances mapped for hot, arid regions and certain coastal environments provide a guide to where the addition of DSS sites to the INDEPTH network would improve environmental coverage most significantly.

In addition to requiring estimates of how confidently DSS findings may be extended to other rural areas without a DSS, information on how the DSS sites within the rural INDEPTH network are interrelated environmentally is required. The dendrogram in Figure 3 provides a valuable map of such environmental similarities, with the main suggested grouping to match the ecozones marked. The groupings of sites linked by lower branches enable further intercomparison.

Class 1 is the least well represented environmentally. A large proportion of its area has Euclidean distances >100 000 (46.4%), compared with just 0.3%, 0.9% and 0.3% for classes

2, 3 and 4 respectively (Figure 4). Class 1 also displays the largest mean Euclidean distance. The class also has the lowest mean distance within the first quartile of distances, showing that the southern Sahel region is environmentally very similar to the current network of DSS sites. Figure 1 shows that the less-vegetated, drier conditions of the northern Sahel, horn of Africa, northern Kenya and southwest Angola are relatively poorly represented by the rural INDEPTH network in Africa, however. Findings and relationships derived from class 1 sites can therefore be extrapolated and applied across the southern Sahel region with confidence, but uncertainty levels increase with latitude.

Class 2 is represented by just Dikgale (South Africa) and Mlomp (Senegal) DSS sites, and while it only covers significant areas of Zimbabwe, Botswana and South Africa, it exhibits a wide range of Euclidean distance values. Eastern Botswana and northern South Africa are most similar environmentally to the conditions found in the DSS sites (Figure 4b). In contrast, central South Africa, Lesotho and Namibia all exhibit relatively large environmental distances from the network. The wide variation in Euclidean distances over such a relatively small area in southern Africa shows that DSS findings should be extended to surrounding regions with caution.

The more hospitable savannah of class 3 covers the largest populated area of Africa and has a low mean Euclidean distance relative to classes 1 and 2, suggesting a relatively good INDEPTH network coverage for these regions. The lowest distance values are across inland western Africa, where the environment is well represented by the Nouna, Farafenni, Navrongo and Bandafassi INDEPTH sites, which span the border between classes 1 and 3. Central Mozambique, Tanzania and Ethiopia are also environmentally similar to class 3 rural DSS sites. Elsewhere, large areas of Mozambique, Angola, Gabon, Congo, the Democratic Republic of the Congo (DRC) and particularly South Africa are highlighted as environmentally most dissimilar to the current INDEPTH network for class 3. It is these areas where the environmental differences contribute to uncertainty in extrapolating findings from the INDEPTH network, and where the addition of new DSS sites would improve the environmental coverage of the network for the savannah zone.

Figure 4d shows that for the cooler, wetter and greener areas of Africa, there is a relatively comprehensive coverage, with a mean Euclidean distance of only 8272 and relatively low distance values across much of the DRC, Cameroon, Liberia, the Côte d'Ivoire and Ghana. However, the northern region of the DRC, Rwanda, Burundi and eastern Madagascar are distant environmentally from the INDEPTH sites, and therefore there is greater uncertainty in applying any relationships found at individual sites to these regions than elsewhere in class 4.

Future directions

Future research should include extension of the analysis to the environmental coverage of the INDEPTH DSS network in Asia or South/Central America. The use of finer spatial resolution environmental indicators, derived from satellite sensors such as MODIS (Tatem *et al.* 2004) or interpolated local meteorological station data, could provide more precise local estimates of network coverage and optimal DSS locations. More generally, integrating such an analysis as this with data on the environmental sensitivity of morbidity and mortality causes at DSS sites would provide an indication of the potential for extrapolation of findings. Similarly, this also aids in defining the existence and importance of non-environmental factors through examination of areas where environmental variables do not describe the observed variance. Finally, such analyses as described in this article should be seen as a generic requirement to assisting the appropriate synthesis and meta-analyses of epidemiological information from wide geographical areas, and providing a baseline from which to extend findings.

Conclusions

The results suggest that the current architecture of the INDEPTH rural network provides a comprehensive coverage of much of the wide range of climates and environments found across SSA, particularly across the southern Sahel and central tropical regions. This in turn infers that findings and relationships derived at DSS sites can be applied elsewhere within the same environmental class with some confidence. Areas where the current INDEPTH network is not representative in terms of environmental conditions do exist, however. The drier, less vegetated regions of the northern Sahel, horn of Africa and northern Kenya are poorly represented, as are other isolated regions in distinct environmental zones. Within these areas, extrapolation of relationships derived from current INDEPTH sites should be undertaken with caution. Finally, examination of the DSS environmental phylogeny suggests logical summaries into distinct groups, which match environmental class membership.

Acknowledgments

Thanks to Dr Carlos Guerra and Dr Abdisalan Noor for comments on the original manuscript. SIH and AJT are funded by a Research Career Development Fellowship from the Wellcome Trust (#06904). RWS is a Wellcome Trust Senior Research Fellow (#058992) and acknowledges the support of the Kenyan Medical Research Institute (KEMRI). Permission for the publication of this article was obtained from the Director, KEMRI.

References

- Balk D, Deichmann U, Yetman G, Pozzi F, Hay SI, Nelson A. Determining global population distribution: methods, applications and data. *Advances in Parasitology*. 2006; 62:120–156.
- Boyd DS, Curran PJ. Using remote sensing to reduce uncertainties in the global carbon budget: the potential of radiation acquired in the middle infrared wavelengths. *Remote Sensing Reviews*. 1998; 16:293–327.
- Campbell, J. *Introduction to Remote Sensing*. The Guildford Press; London: 1996.
- ERDAS. *ERDAS Field Guide*. 7th Edn.. ERDAS Inc.; Atlanta: 2003.
- FAO. *Global Agro-Ecological Zones*. Food and Agriculture Organization (FAO); Rome: 2000.
- Gallup JL, Sachs JD. The economic burden of malaria. *The American Journal of Tropical Medicine and Hygiene*. 2001; 64:85–96. [PubMed: 11425181]
- Hay SI. An overview of remote sensing and geodesy for epidemiology and public health application. *Advances in Parasitology*. 2000; 47:1–35. [PubMed: 10997203]
- Hay SI, Lennon JJ. Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Tropical Medicine and International Health*. 1999; 4:58–71. [PubMed: 10203175]
- Hay SI, Guerra CA, Tatem AJ, Atkinson PM, Snow RW. Urbanization, malaria transmission and disease burden in Africa. *Nature Reviews Microbiology*. 2005a; 3:81–90.
- Hay SI, Noor AM, Nelson A, Tatem AJ. The accuracy of human population maps for public health application. *Tropical Medicine and International Health*. 2005b; 10:1073–1086. [PubMed: 16185243]
- Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ. Global environmental data for mapping infectious disease distributions. *Advances in Parasitology*. 2006; 62:38–77.
- INDEPTH. *Population, Health and Survival at INDEPTH Sites*. Vol. 1. International Development Research Centre; Ottawa: 2002. Population and health in developing countries.
- INDEPTH. *Measuring Health Equity in Small Areas: Findings from Demographic Surveillance Systems*. Ashgate Publishing Ltd, Aldershot; 2005.
- Murray CJL, Lopez AD. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *The Lancet*. 1997; 349:1269–1276.
- Ngom P, Binka FN, Phillips JF, Pence B, Macleod B. Demographic surveillance and health equity in sub-Saharan Africa. *Health Policy and Planning*. 2001; 16:337–344. [PubMed: 11739357]

- Olson DM, Dinerstein E, Wikramanayake ED, et al. Terrestrial ecoregions of the World: a new map of life on Earth. *BioScience*. 2001; 51:933–938.
- Price JC. Land surface temperature measurements from the split window channels of the NOAA 7 advanced very high resolution radiometer. *Journal of Geophysical Research*. 1984; 89:7231–7237.
- Rogers DJ. Satellites, space, time and the African trypanosomiases. *Advances in Parasitology*. 2000; 47:129–171. [PubMed: 10997206]
- Rogers, DJ.; Robinson, TP. Tsetse distribution. In: Maudlin, I.; Holmes, PH.; Miles, MA., editors. *The Trypanosomiases*. CABI Publishing; Wallingford: 2004. p. 139-180.
- Rogers DJ, Randolph SE, Snow RW, Hay SI. Satellite imagery in the study and forecast of malaria. *Nature*. 2002; 415:710–715. [PubMed: 11832960]
- Schultz, J. *The Ecozones of the World*. Springer-Verlag; Berlin: 2005.
- Strahler, AH.; Strahler, A. The Koeppen climate system in physical geography. In: Strahler, AH.; Strahler, A., editors. *Physical Geography: Science and Systems of the Human Environment*. John Wiley & Sons; Chichester: 2002. p. 244-247.
- Sugihara G, Bersier L-F, Southwood TRE, Pimm SL, May RM. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proceedings of the National Academy of Sciences of the USA*. 2003; 100:5246–5251. [PubMed: 12702773]
- Tatem AJ, Hay SI. Measuring urbanization pattern and extent for malaria research: a review of remote sensing approaches. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*. 2004; 81:363–376. [PubMed: 15273262]
- Tatem AJ, Goetz SJ, Hay SI. Terra and Aqua: new data for epidemiology and public health. *Journal of Applied Earth Observation and Geoinformation*. 2004; 6:33–46.
- Tatem AJ, Hay SI, Rogers DJ. Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences of the USA*. 2006a; 103:6242–6247. [PubMed: 16606847]
- Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. *Advances in Parasitology*. 2006b; 62:293–343. [PubMed: 16647974]
- Webb, A. *Statistical Pattern Recognition*. Wiley; Chichester: 2003.

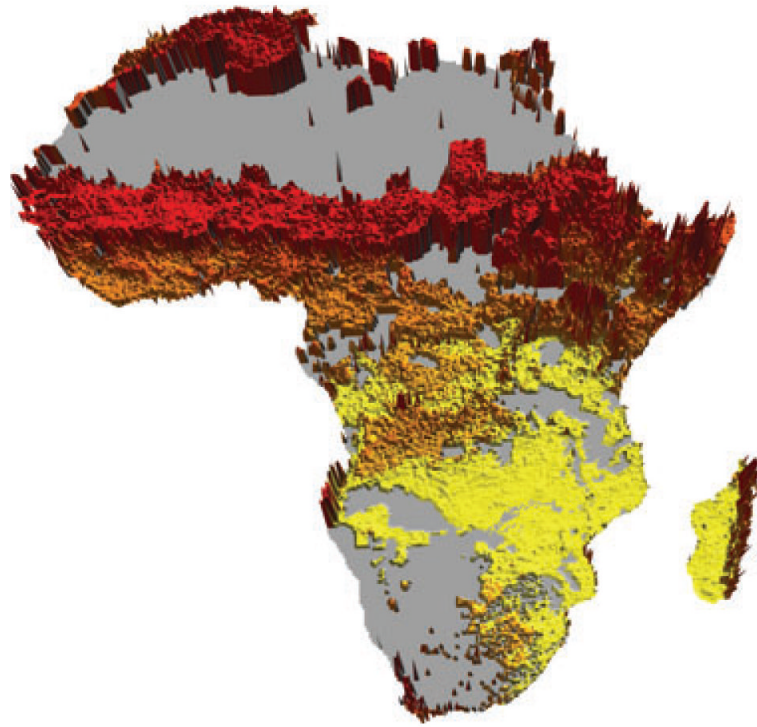


Figure 1. Environmental distance from the INDEPTH DSS network in Africa. Yellow represents the lowest through red the highest values with a vertical exaggeration applied to emphasise the differences. Areas with population densities of <1 person per km^2 have been masked in grey.

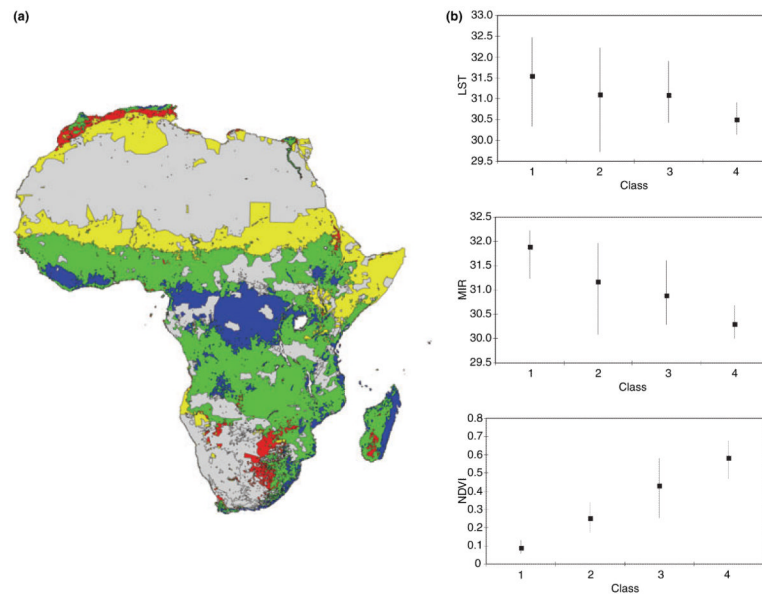


Figure 2. (a) Environmental classes produced through ISODATA clustering and dendrogram analysis. Yellow, class 1; red, class 2; green, class 3; blue, class 4. Areas with population densities of <math><1</math> person per

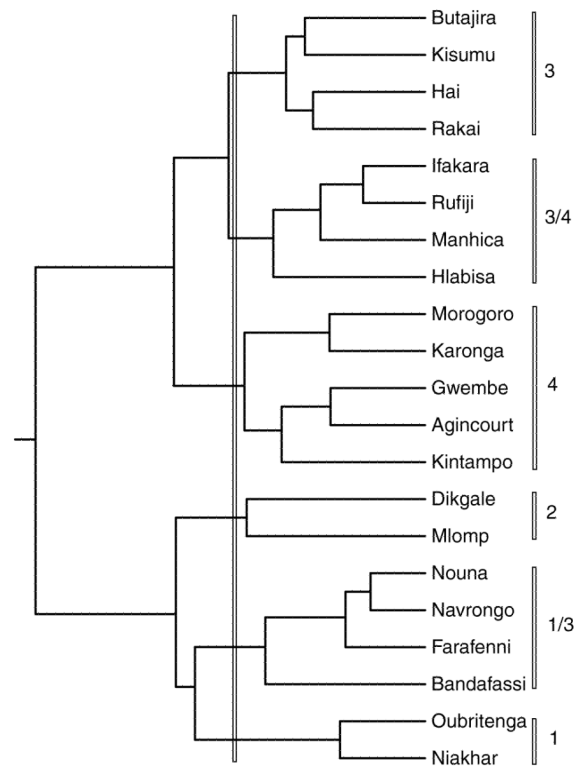


Figure 3. Environmental dendrogram of the INDEPTH DSS sites in Africa, with the environmental class of each and grouping cut-off matches these classes marked.

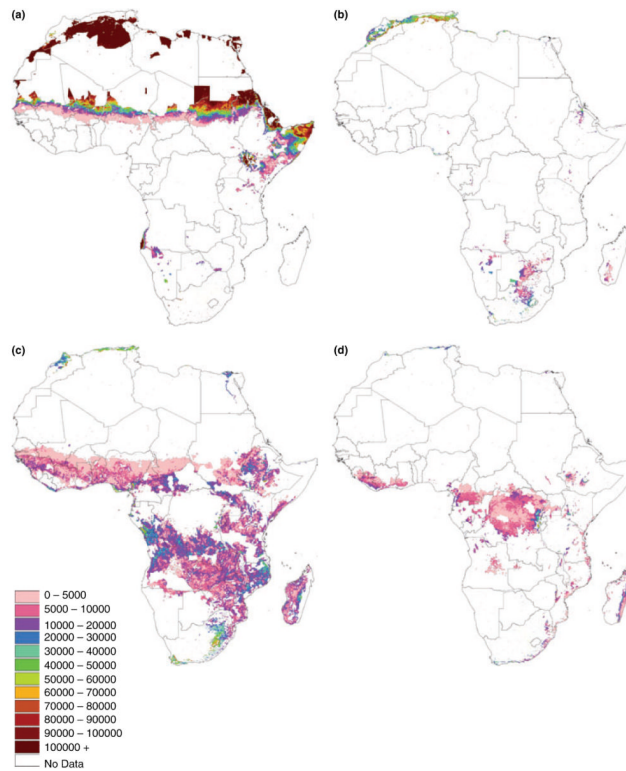


Figure 4.

Environmental distance images for (a) class 1, (b) class 2, (c) class 3 and (d) class 4. The mean Euclidean distance for (a) is 65 387, (b) is 27 134, (c) is 11 551 and (d) is 8272. A large Euclidean distance indicates a relatively different environment from that of the INDEPTH DSS sites, while a small distance represents a relatively similar environment.