Supplementary file 1. Generating alternative classes of urbanicity by DHS cluster and modelling travel time to health facilities

As an alternative to the DHS urban and rural classifications, we derived three classes of the urban continuum (urbanicity) - rural, semi-urban, and core urban based on satellite imagery. We used the 2015 Global Human Settlement Layer- settlement model (GHS-SMOD)[1][2] to classify the location of DHS clusters into different degrees of urbanicity. GHS-SMOD delineates and classifies settlement typologies through cell clusters' population size, population, and built-up area densities based on the Built-up (GHS-BUILT) areas and Population (GHS-POP) data layers. GHS-BUILT represents the physical extent of the human settlement produced through automatic supervised classification of the Landsat and Sentinel satellite imagery while GHS-POP, a high spatial resolution (250×250 m² and 1×1 km²) population density layer is produced by downscaling national census counts data at district level to a regular fine scale grid. It is the combination of GHS-BUILT and GHS-POP based on the *degree of urbanization* concept[3] that results in GHS-SMOD at the 1 km spatial resolution.[4]

Each pixel of the utilized GHS-SMOD layer[1] contained a single urbanicity class based on local population density, permanency of the water body, 30% or 50% of a pixel being built-up surface and generalization through smoothing and gap filling.[2] Based on these rules, level 1 encapsulates 3 classes (urban centre, urban cluster, and rural grid cell) which are further broken down (level 2) into seven classes that were used in this analysis namely; urban centre (class 30), dense urban cluster (class 23), semi-dense urban cluster (class 22), sub-urban or peri-urban (class 21), rural cluster (class 13), low density rural (class 12), very low-density rural (class 11), and water (class 10) grid cells. A detailed description of how these data are generated, processed, and classified is available elsewhere.[2][5] Supplementary file 1 shows the spatial distribution of the seven classes based on the 2015 GHS-SMOD layer in Tanzania describing the continuum from urban to rural areas in 2015. Urbanicity classes were first reclassified to an ordinal scale from 1 (least urban) to 7 (most urban) after masking out the water class. The average, majority, maximum and minimum values were extracted per buffer and used to define three new classes of urbanicity that were used for the bivariate and multi-level multivariate logistic regression analysis. Buffers were used to reduce the bias associated with scrambling cluster coordinates. We created 2km (urban clusters) and 5km (rural clusters) circular buffers and extracted the polygonal properties of urbanicity as previously implemented[6][7] and recommended.[8]

Due to the low count of neonatal deaths and homogeneity in some of the extracted urbanicity classes, the seven classes were collapsed into three classes based on the following criteria. Class 1 (core urban) where the mean of all cells per buffer was at least 6, and the maximum and majority of cells per buffer; Class 2 (semi-urban and areas in transition): The mean per buffer is 3 or above but less than 6, or the mean value was less than 3 but had a maximum of at least 4 or above to account for small elements of urban areas such as a small town surrounded by rural areas; the rest of the clusters were assigned to Class 3 (rural areas) that is, where all means were less than or equal to 2 while the maximum values per buffer were at least 4.

Modelling travel time to hospitals

Given that short distances in urban areas can obscure long travel times,[9] we also included a consideration for accessibility of emergency obstetric healthcare during pregnancy and childbirth generally provided only in hospitals as a potential explanation (effect moderator) between urbanicity and neonatal mortality. A proxy of geographic accessibility to hospital was not available in the DHS and was thus modelled independently for each cluster. It was proxied by the time taken to travel between a DHS cluster and the nearest public hospital, based on a least-cost path algorithm implemented in a Geographic Information System (GIS) via WHO AccessMod 5 software (alpha version 5.7.8)[10] widely used across healthcare applications in SSA.[11] We first assembled spatial layers of factors that affect travel time which included ESA Sentinel-2 landcover at 10m x 10m spatial resolution,[12] road network from OpenStreetMaps (OSM), NASA Shuttle Radar Topography Mission (SRTM) digital elevation model at 30m x 30m spatial resolution, water bodies and protected areas.[13] The land cover had nine classes (water, flooded vegetation, trees, ice/snow, grass, shrubs, crops, built-up areas and bare ground), while roads were re-classified into four classes (primary, secondary, tertiary, and minor roads) based on OSM description.

The spatial layers were resampled to 300m and merged to form a single layer via the *Accessibility module* in AccessMod software. Travel speeds were then applied on the merged layer to generate cumulative travel time from each cell (pixel) to the nearest hospital in mainland Tanzania at 300 x 300m spatial resolution. Two modes of transport were considered, walking while travelling on off-roads cells,

and driving on motorable roads. The adopted modes and travel speeds across the different land cover and road classes were informed by previous studies in similar contexts.[14-16] Further, walking speeds were corrected for slope derived from DEM using Tobler's hiking function, an exponential function that describes how human walking speed varies with slope.[17] The base layer of hospitals was derived from a geolocated pan African master health facility list of public health service providers.[18] After verification, the list contained 236 public hospitals in Tanzania. The result was a gridded dataset showing travel time to the nearest public hospital in 2015 at 300m spatial resolution. We then linked each DHS cluster with its corresponding travel time to the nearest hospital. Similar to urbanicity, we extracted the average travel time as a continuous variable. For three clusters located in Maisome, Ikuza and Bulyalike islands we used reported travel times from co-author familiar with these regions (ABP). Modelled travel time was not available for these locations because we did not incorporate water as means of transport due to lack of data to parametrise the model.

Table S1: The mean travel time to the nearest hospital from each cluster stratified based on DHS (urban and rural) and GHS-SMOD (core urban, semi-urban and rural) classifications in Tanzania. The DHS and GHS-SMOD represents the same spatial extents

Source	Urban/rural classification	Number of clusters	Mean travel time (minutes)
DHS	Urban	163	14.4
	Rural	364	78.2
GHS-SMOD	Core urban	61	4.3
	Semi-urban	224	41.1
	Rural	242	89.4
National 8915		527	62.8

Figure S1. The spatial distribution of the seven classes based on the 2015 GHS-SMOD layer in Tanzania describing the continuum from urban to rural areas in 2015.

Urbanization

- Very low density rural (1)
- Low Density Rural (2)
- Rural cluster (3)
- Suburban or per-urban (4)
- Semi-dense Urban Cluster (5)
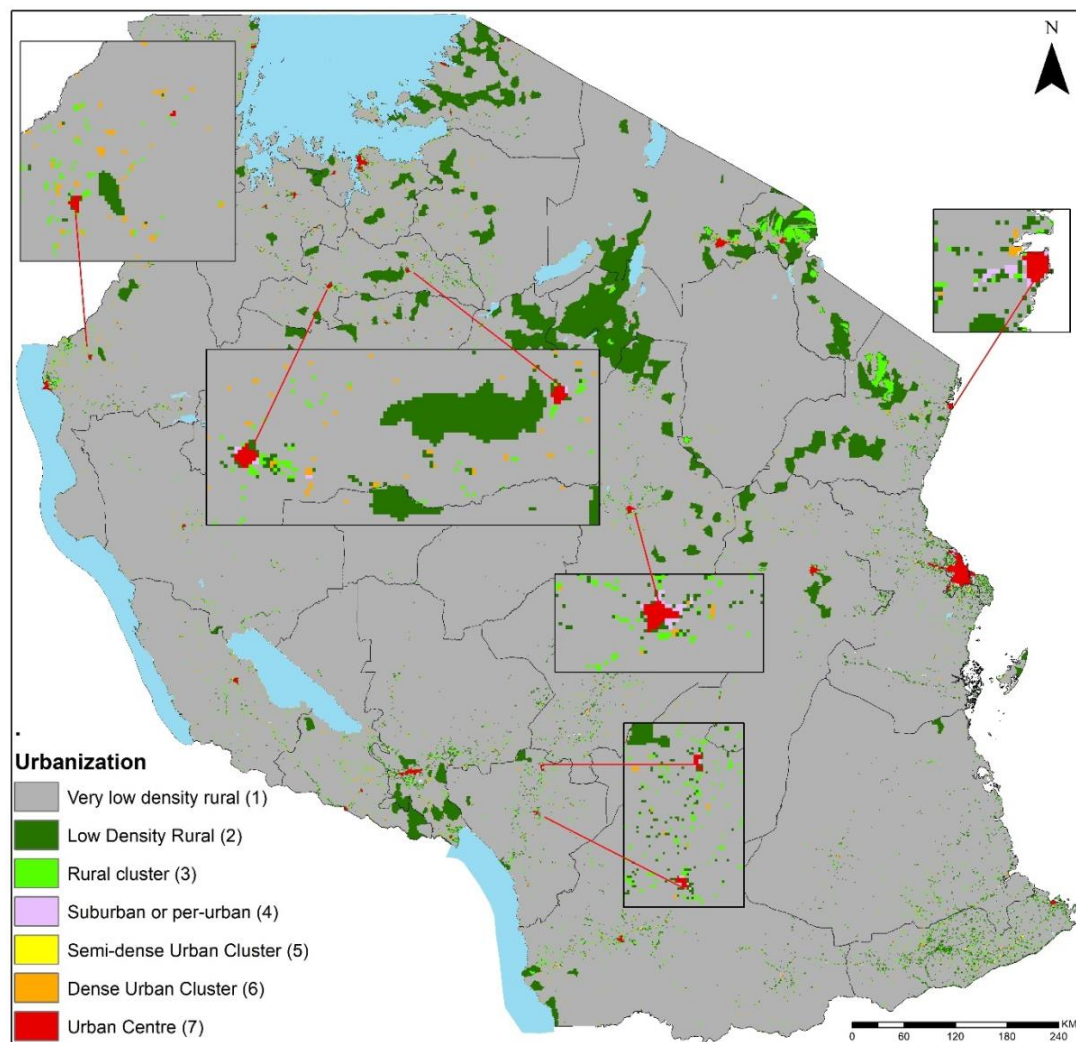- Dense Urban Cluster (6)
- Urban Centre (7)

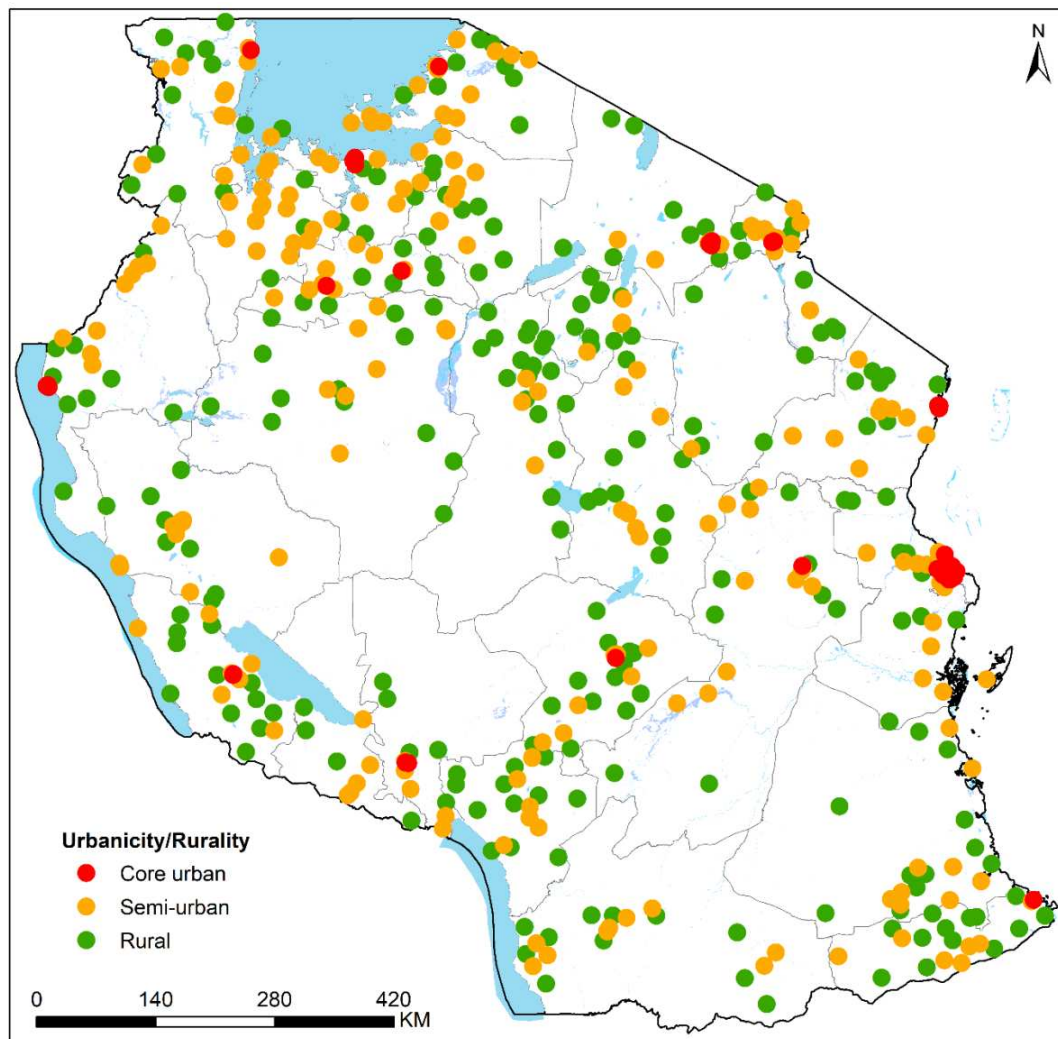Figure S2. GHS-SMOD-derived urbanicity classifications based on satellite imagery

Figure S3. Travel time to the nearest hospital (n=236) in Tanzania in 2015 at 300m spatial resolution classified into 5 classes ranging from less than 30 minutes (green) to over 2 hours/120 minutes (red). The white areas are national parks/reserves that were considered as barriers except in presence of roads.
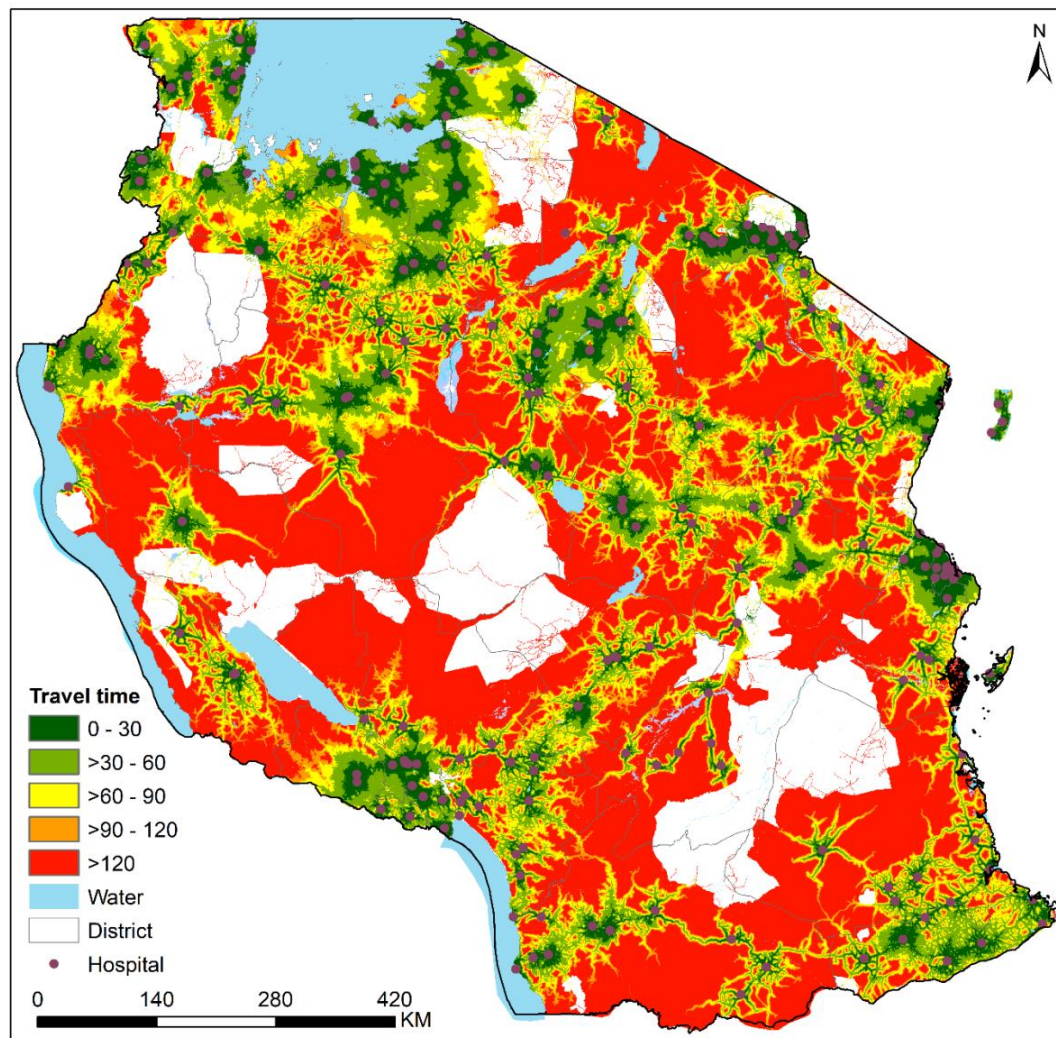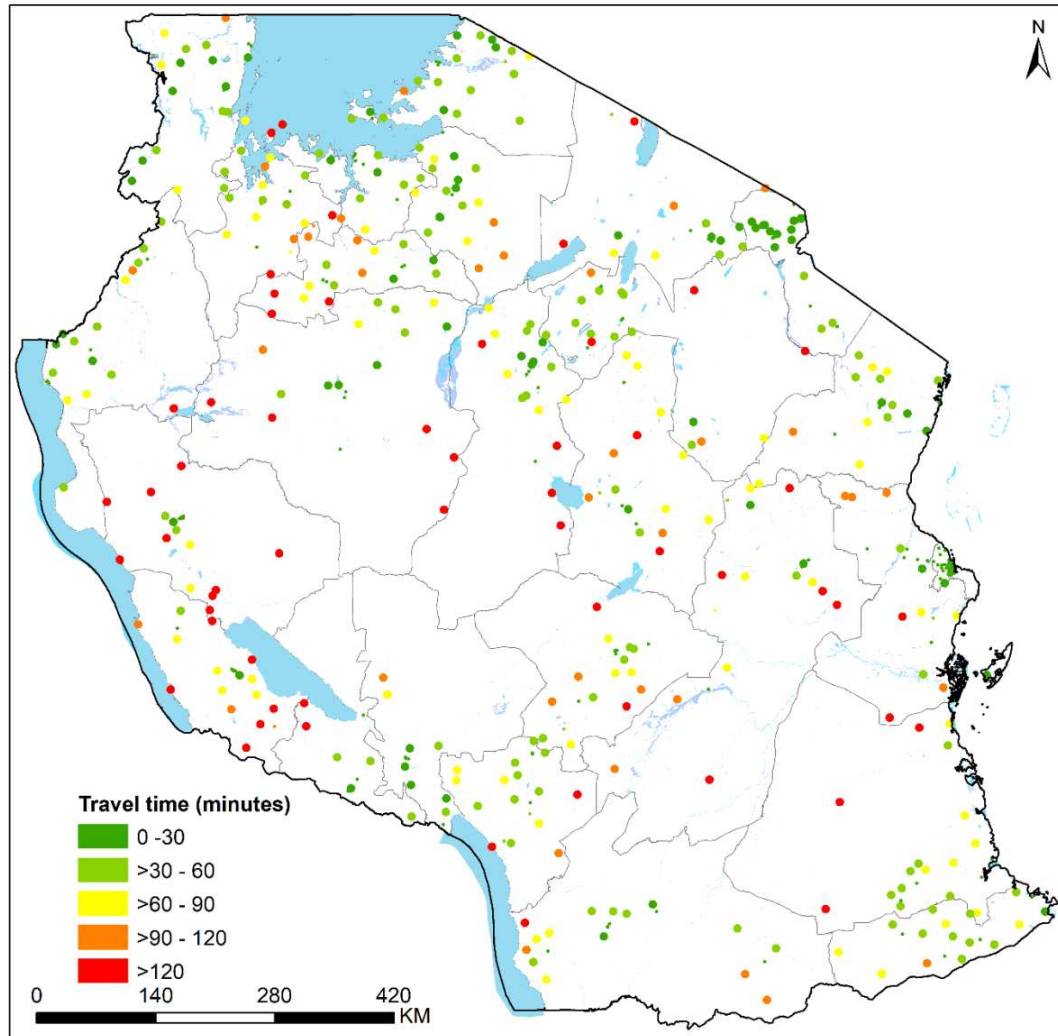
Figure S4. Travel time to the nearest hospital (n=236) in Tanzania in 2015 at cluster level classified into five classes ranging from less than 30 minutes (green) to +120 minutes (red)

References

1. Pesaresi M, Florczyk A, Schiavina M, et al. GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A 2019 [Available from: http://data.europa.eu/89h/42e8be89-54ff-464e-be7b-bf9e64da5218

2. Florczyk A, Corbane C, Ehrlich D, et al. GHSL Data Package 2019, EUR 29788 EN Luxembourg [Available from: https://publications.jrc.ec.europa.eu/repository/handle/JRC117104 accessed September 2022.

3. Dijkstra L, Poelman H. A harmonised definition of cities and rural areas: the new degree of urbanisation 2014 [Available from: http://ec.europa.eu/regional_policy/en/information/publications/working-papers/2014/a-harmonised-definition-of-cities-and-rural-areas-the-new-degree-of-urbanisation.

4. Florczyk A, Melchiorri M, Corbane C, et al. Description of the GHS Urban Centre Database 2015, Public Release 2019, Version 1.0, Luxembourg2019 [Available from: https://publications.jrc.ec.europa.eu/repository/handle/JRC115586 accessed October 2022.

5. Freire S, Corbane C, Zanchetta L, et al. GHSL data package 2019 : public release GHS P2019: Publications Office 2019.

6. Burgert CR, Colston J, Roy T, et al. Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys. DHS Spatial Analysis Reports No. 7. Calverton, Maryland, USA: ICF International 2013.

7. Ruktanonchai CW, Ruktanonchai NW, Nove A, et al. Equality in maternal and newborn health: modelling geographic disparities in utilisation of care in five East African countries. *PLoS One* 2016;11(8):e0162006.

8. Masters SH, Burstein R, Amofah G, et al. Travel time to maternity care and its effect on utilization in rural Ghana: a multilevel analysis. *Social Science & Medicine* 2013;93:147-54.

9. Banke-Thomas A, Wong KLM, Collins L, et al. An assessment of geographical access and factors influencing travel time to emergency obstetric care in the urban state of Lagos, Nigeria. *Health Policy and Planning* 2021;36(9):1384-96. doi: 10.1093/heapol/czab099

10. Ray N, Ebener S. AccessMod 3.0: computing geographic coverage and accessibility to health care services using anisotropic movement of patients. *International journal of health geographics* 2008;7(1):1-17.

11. Ouma PO, Maina J, Thuranira PN, et al. Access to emergency hospital care provided by the public sector in sub-Saharan Africa in 2015: a geocoded inventory and spatial analysis. *The Lancet Global Health* 2018;6(3):e342-e50.

12. Zanaga D, Van De Kerchove R, De Keersmaecker W, et al. ESA WorldCover 10 m 2020 v100 (Version v100) 2021 [Available from: https://doi.org/10.5281/zenodo.5571936.

13. UNEP-World Conservation Monitoring Centre (WCMC), International Union for Conservation of Nature (IUCN). The world database on protected areas 2017 [Available from: https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA accessed 23 July 2020.

14. Fogliati P, Straneo M, Brogi C, et al. How Can Childbirth Care for the Rural Poor Be Improved? A Contribution from Spatial Modelling in Rural Tanzania. *PLOS ONE* 2015;10(9):e0139460. doi: 10.1371/journal.pone.0139460

15. Pozzi F, Robinson TP. Accessibility mapping in the Horn of Africa: Applications for livestock policy. *IGAD LPI Working Paper 11-08* 2008

16. Macharia PM, Odera PA, Snow RW, et al. Spatial models for the rational allocation of routinely distributed bed nets to public health facilities in Western Kenya. *Malaria Journal* 2017;16(1):367. doi: 10.1186/s12936-017-2009-3

17. Tobler W. Three presentations on geographical analysis and modeling. University of California, 1993.

18. Maina J, Ouma PO, Macharia PM, et al. A spatial database of health facilities managed by the public health sector in sub Saharan Africa. *Scientific Data* 2019;6(1):134. doi: 10.1038/s41597-019-0142-2