

## SUPPLEMENTARY MATERIAL

### **Aggregating GlobCover land cover classes for dasymetric population modelling**

The regional GlobCover dataset includes 47 different land cover classes. This very detailed classification does not necessarily add useful information for population modelling. A high number of different land cover classes would only improve the accuracy of population distribution data produced if per-land cover class population densities are significantly different. The optimal land cover data for population distribution modelling would be a classification that maximizes within land cover class homogeneity and maximizes between-class heterogeneity in relation to population density. Here we tested the effect of aggregating GlobCover land cover classes on the accuracy of the output population distribution datasets produced for Kenya.

#### **1. Method**

GlobCover is compatible with the UN Land Cover Classification System (LCCS). This allows easy aggregation of LC classes based on a hierarchy of LC class detail, and allows comparison of LC classes across countries and regions. We tested two levels of land cover class aggregation. The 47 individual classes were first aggregated to 23 classes and secondly, to 10 generic classes (table S1).

We followed the same procedure as described in the main document and in figure 1. For this analysis, we fixed EXL to L1 and varied TCM. The procedure was repeated using GlobCover with three different levels of land cover class aggregation: no aggregation (47 classes), level 1 (23 classes) and level 2 (10 classes). Each stage was replicated 25 times. Root mean square errors (RMSE) between observed and predicted population data per EA were compared. An analysis of variance was performed to test for the impact of different numbers of land cover classes on the RMSE.

#### **2. Results**

Figure S1 shows that the accuracy of population data produced using the aggregation level 1 GlobCover dataset (23 classes) and the original GlobCover dataset (47 classes) were very close. The aggregation level 2 had an opposite effect on population distribution accuracies, with a lower average RMSE at the ADMIN-2 level, and a higher average RMSE at the ADMIN-1 level. For the other levels of TCM, the RMSEs were very close, whichever aggregation level was used. An analysis of variance including the aggregation level of GlobCover data and the TCM as independent variables confirmed that the aggregation level used had no significant impact on the RMSE (F value = 0.17; p = 0.84).

### **3. Conclusion**

In the main document, analyses showed that GlobCover, combined with detailed settlement extents, generally provided significantly more accurate population distribution models than other global land cover datasets in Kenya. The present additional analysis shows that combining GlobCover land cover classes does not decrease the accuracy of population distribution models for Kenya. Using aggregated land cover data to simplify the population modelling procedure enables the calculation of more accurate population weights without degrading the quality of the output population datasets produced.

**Table S1:** GlobCover land cover classification with two levels of class aggregation

| Value | GlobCover regional classes                                                            | Aggregation level 1                                                         | Aggregation level 2                                             |
|-------|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|-----------------------------------------------------------------|
| 11    | Post-flooding or irrigated croplands (or aquatic)                                     | Post-flooding or irrigated croplands (or aquatic)                           | Cultivated Terrestrial Areas and Managed Lands                  |
| 12    | Post-flooding or irrigated shrub or tree crops                                        |                                                                             |                                                                 |
| 13    | Post-flooding or irrigated herbaceous crops                                           |                                                                             |                                                                 |
| 14    | Rainfed croplands                                                                     | Rainfed croplands                                                           |                                                                 |
| 15    | Rainfed herbaceous crops                                                              |                                                                             |                                                                 |
| 16    | Rainfed shrub or tree crops (cash crops, vineyards, olive tree, orchards...)          |                                                                             |                                                                 |
| 20    | Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)           | Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%) |                                                                 |
| 21    | Mosaic cropland (50-70%) / grassland or shrubland (20-50%)                            |                                                                             |                                                                 |
| 30    | Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)           | Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%) |                                                                 |
| 31    | Mosaic grassland or shrubland (50-70%) / cropland (20-50%)                            |                                                                             |                                                                 |
| 32    | Mosaic forest (50-70%) / cropland (20-50%)                                            |                                                                             |                                                                 |
| 40    | Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)            | Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)  | Natural and Semi-natural Terrestrial Vegetation – Woody / Trees |
| 41    | Closed (>40%) broadleaved evergreen and/or semi-deciduous forest (>5m)                |                                                                             |                                                                 |
| 42    | Open (15-40%) broadleaved semi-deciduous and/or evergreen forest with emergents (>5m) |                                                                             |                                                                 |
| 50    | Closed (>40%) broadleaved deciduous forest (>5m)                                      | Closed (>40%) broadleaved deciduous forest (>5m)                            |                                                                 |
| 60    | Open (15-40%) broadleaved deciduous forest/woodland (>5m)                             | Open (15-40%) broadleaved deciduous forest/woodland (>5m)                   |                                                                 |
| 70    | Closed (>40%) needleleaved evergreen forest (>5m)                                     | Closed (>40%) needleleaved evergreen forest (>5m)                           |                                                                 |
| 90    | Open (15-40%) needleleaved deciduous or evergreen forest (>5m)                        | Open (15-40%) needleleaved deciduous or evergreen forest (>5m)              |                                                                 |
| 91    | Open (15-40%) needleleaved deciduous forest (>5m)                                     |                                                                             |                                                                 |
| 92    | Open (15-40%) needleleaved evergreen forest (>5m)                                     |                                                                             |                                                                 |
| 100   | Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)                 | Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)       |                                                                 |
| 101   | Closed (>40%) mixed broadleaved and needleleaved forest (>5m)                         |                                                                             |                                                                 |
| 110   | Mosaic forest or shrubland (50-70%) / grassland (20-50%)                              | Mosaic forest or shrubland (50-70%) / grassland (20-50%)                    |                                                                 |
| 120   | Mosaic grassland (50-70%) / forest or shrubland (20-50%)                              | Mosaic grassland (50-70%) / forest or shrubland (20-50%)                    |                                                                 |

|     |                                                                                                                                |                                                                                                                                |                                                             |
|-----|--------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|
| 130 | Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)                                    | Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)                                    | Natural and Seminatural Terrestrial Vegetation – Shrubs     |
| 131 | Closed to open (>15%) broadleaved or needle-leaved evergreen shrubland (<5m)                                                   |                                                                                                                                |                                                             |
| 134 | Closed to open (>15%) broadleaved deciduous shrubland (<5m)                                                                    |                                                                                                                                |                                                             |
| 140 | Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)                                            | Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)                                            | Natural and Seminatural Terrestrial Vegetation – Herbaceous |
| 141 | Closed (>40%) grassland                                                                                                        |                                                                                                                                |                                                             |
| 150 | Sparse (<15%) vegetation                                                                                                       | Sparse (<15%) vegetation                                                                                                       | Natural and Semi-natural Terrestrial Vegetation             |
| 151 | Sparse (<15%) grassland                                                                                                        |                                                                                                                                |                                                             |
| 152 | Sparse (<15%) shrubland                                                                                                        |                                                                                                                                |                                                             |
| 160 | Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily) - Fresh or brackish water         | Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily) - Fresh or brackish water         | Natural and Seminatural Aquatic Vegetation                  |
| 161 | Closed to open broadleaved forest on (semi-)permanently flooded land - Fresh water                                             |                                                                                                                                |                                                             |
| 162 | Closed to open broadleaved forest on temporarily flooded land - Fresh water                                                    |                                                                                                                                |                                                             |
| 170 | Closed (>40%) broadleaved forest or shrubland permanently flooded - Saline or brackish water                                   | Closed (>40%) broadleaved forest or shrubland permanently flooded - Saline or brackish water                                   |                                                             |
| 180 | Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil - Fresh, brackish or saline water | Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil - Fresh, brackish or saline water |                                                             |
| 181 | Closed to open (>15%) woody vegetation on regularly flooded or waterlogged soil - Fresh or brackish water                      |                                                                                                                                |                                                             |
| 185 | Closed to open (>15%) grassland on regularly flooded or waterlogged soil - Fresh or brackish water                             |                                                                                                                                |                                                             |
| 190 | Artificial surfaces and associated areas (Urban areas >50%)                                                                    | Artificial surfaces and associated areas (Urban areas >50%)                                                                    | Artificial Surfaces                                         |
| 200 | Bare areas                                                                                                                     | Bare areas                                                                                                                     | Bare areas                                                  |
| 201 | Consolidated bare areas (hardpans, gravels, bare rock, stones, boulders)                                                       |                                                                                                                                |                                                             |
| 202 | Non-consolidated bare areas (sandy desert)                                                                                     |                                                                                                                                |                                                             |
| 203 | Salt hardpans                                                                                                                  |                                                                                                                                |                                                             |
| 210 | Water bodies                                                                                                                   | Water bodies                                                                                                                   | Inland Waterbodies, snow and ice                            |
| 220 | Permanent snow and ice                                                                                                         | Permanent snow and ice                                                                                                         |                                                             |
| 230 | No data (burnt areas, clouds,...)                                                                                              | No data (burnt areas, clouds,...)                                                                                              | No data                                                     |

**Figure S1:** Results from accuracy assessments of population distribution data produced with EXL = L1. Boxplots show the RMSEs according to the TCM and the number of GlobCover land cover classes used as input data. Each stage was replicated 25 times. The dotted line corresponds to the RMSE associated with the areal weighted method (i.e. homogenous distribution of people within administrative units) for each administrative level.

