*Supplementary Information S1. Random Forest Variable Importance by Country and Model*

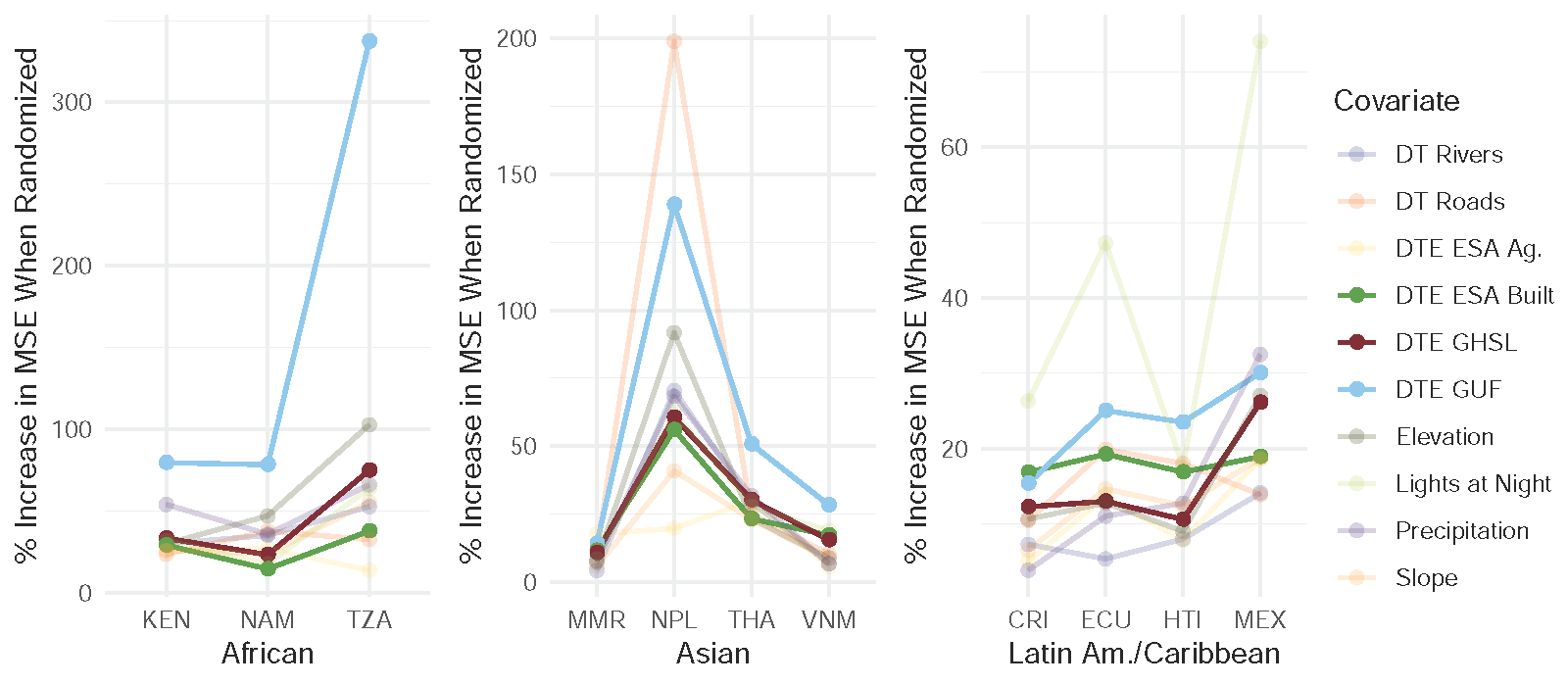


Fig. S1.1. Variable importance plot for the GUF+GHSL Random Forest. The top ten covariates in terms of variable importance as ranked by percentage increase in RMSE when that variable was randomly reassigned to out-of-bag data. These scores are grouped by country and separated by region. DT and DTE refer to distance-to and distance-to-edge respectively. In bold are the three built land cover representations included in the GUF+GHSL RF model.

*Supplementary Information S2. Comparisons of Validation Unit Densities by Country*

These plots compare estimations with observed population density in validation units by model type for each case study country as presented in Fig. 7. Each panel shows the log population density in people/km2 for two thirds of census units used in validation (see Section 2.4 for model acronym explanations). Because of the log scale zeroes are not included on the charts though they affect density percentiles, point, and bootstrapped statistic estimates (Fig. 8, Table 3, S3).

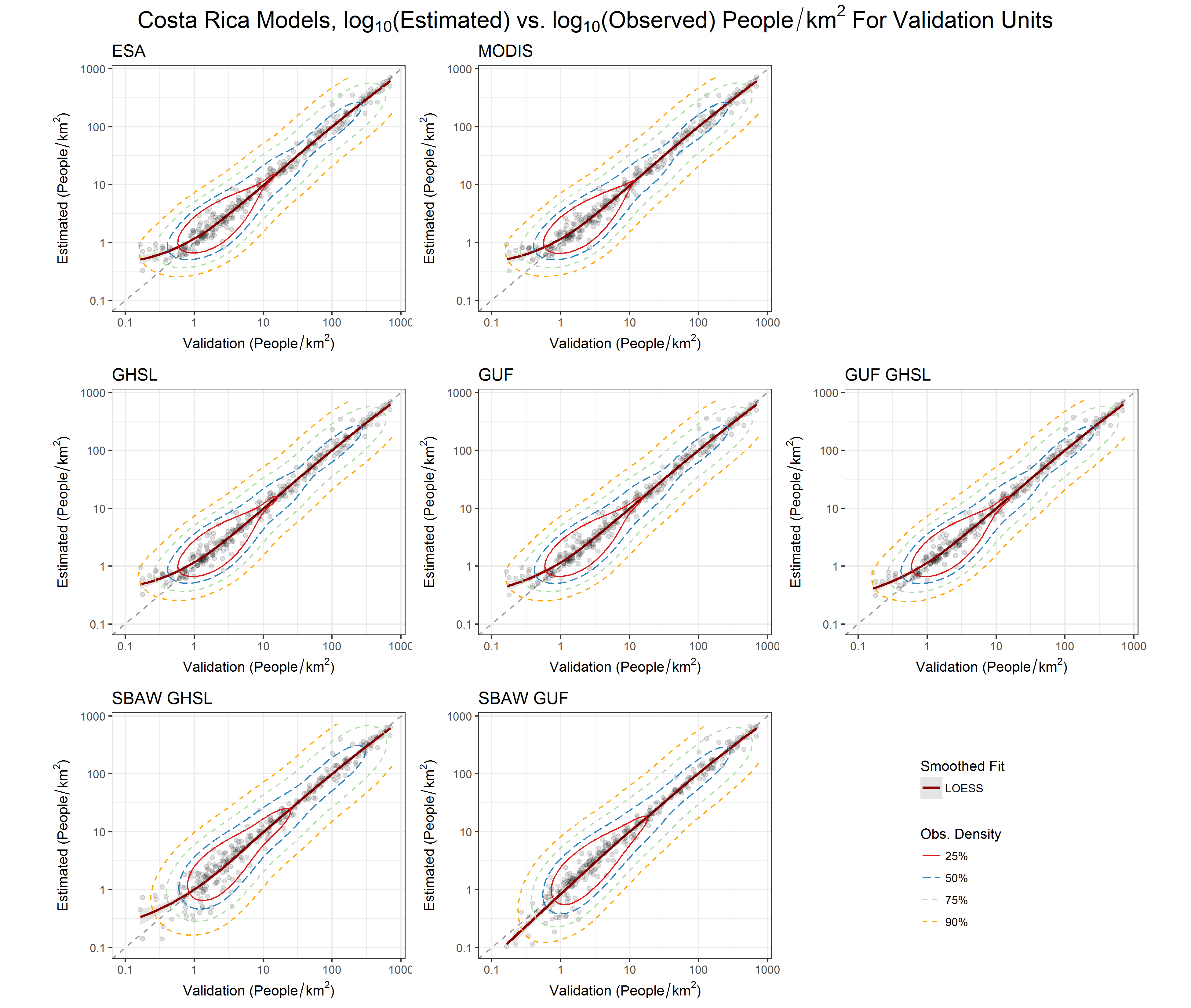


Figure S2.1.

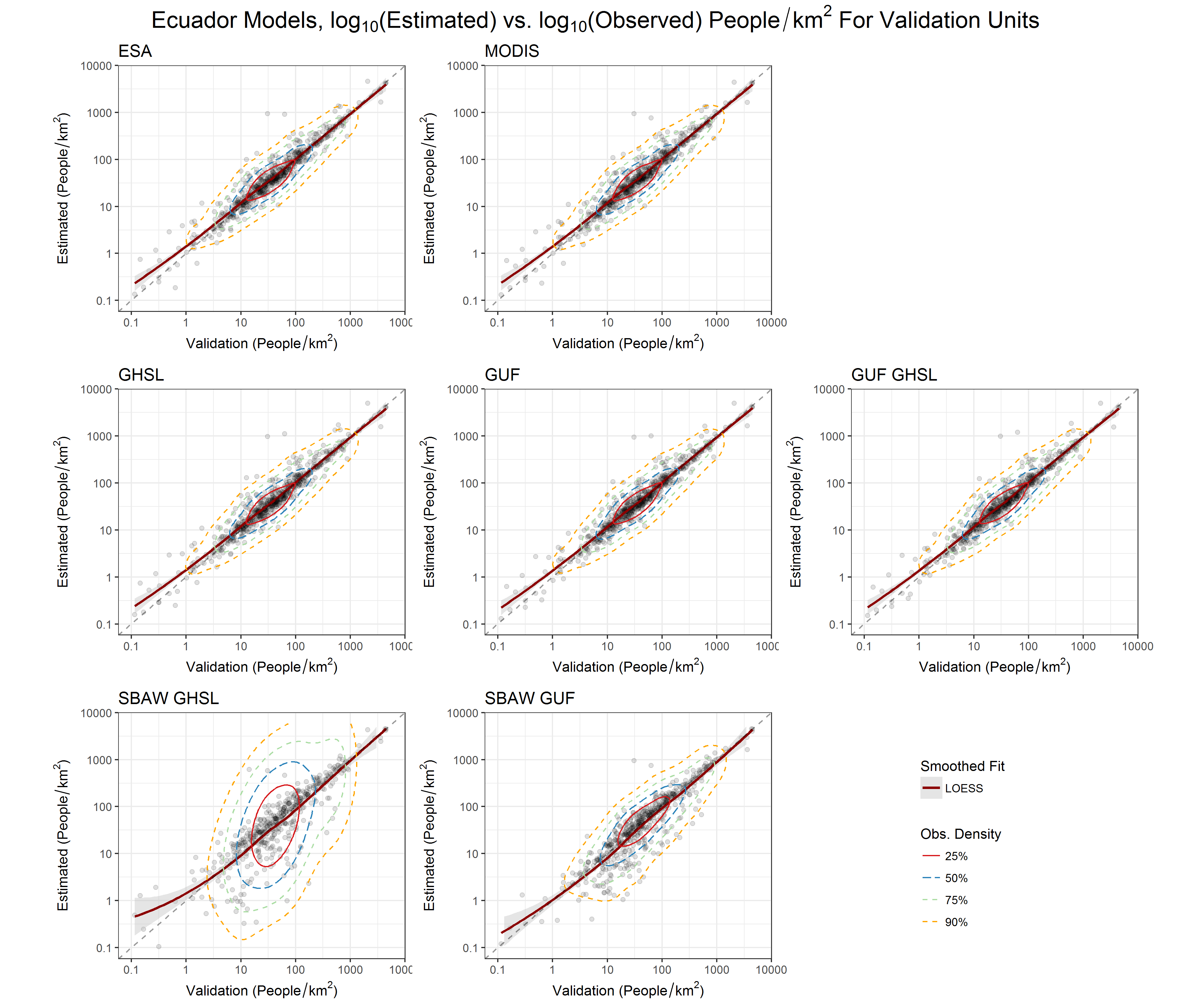


Figure S2.2.

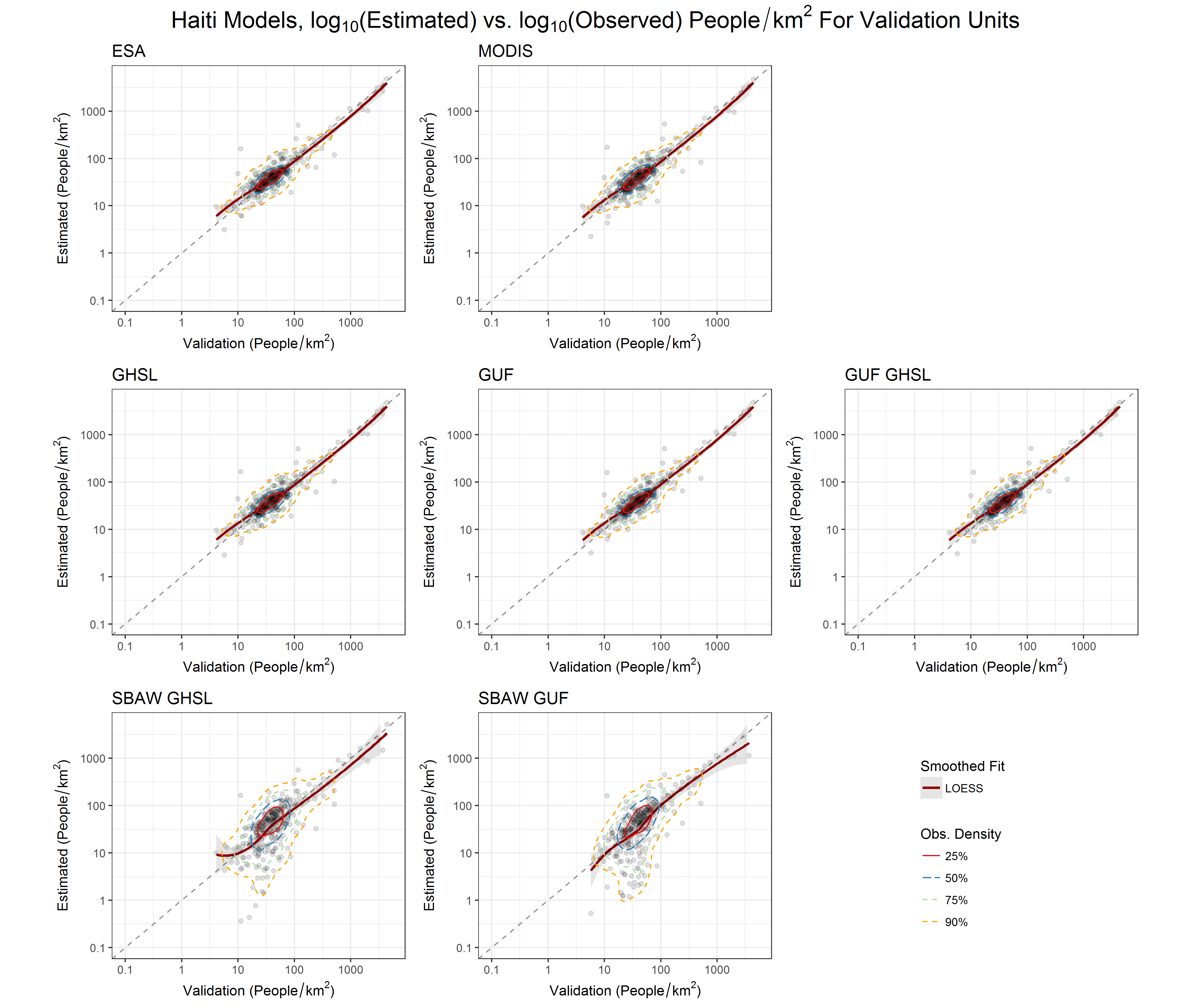


Figure S2.3.

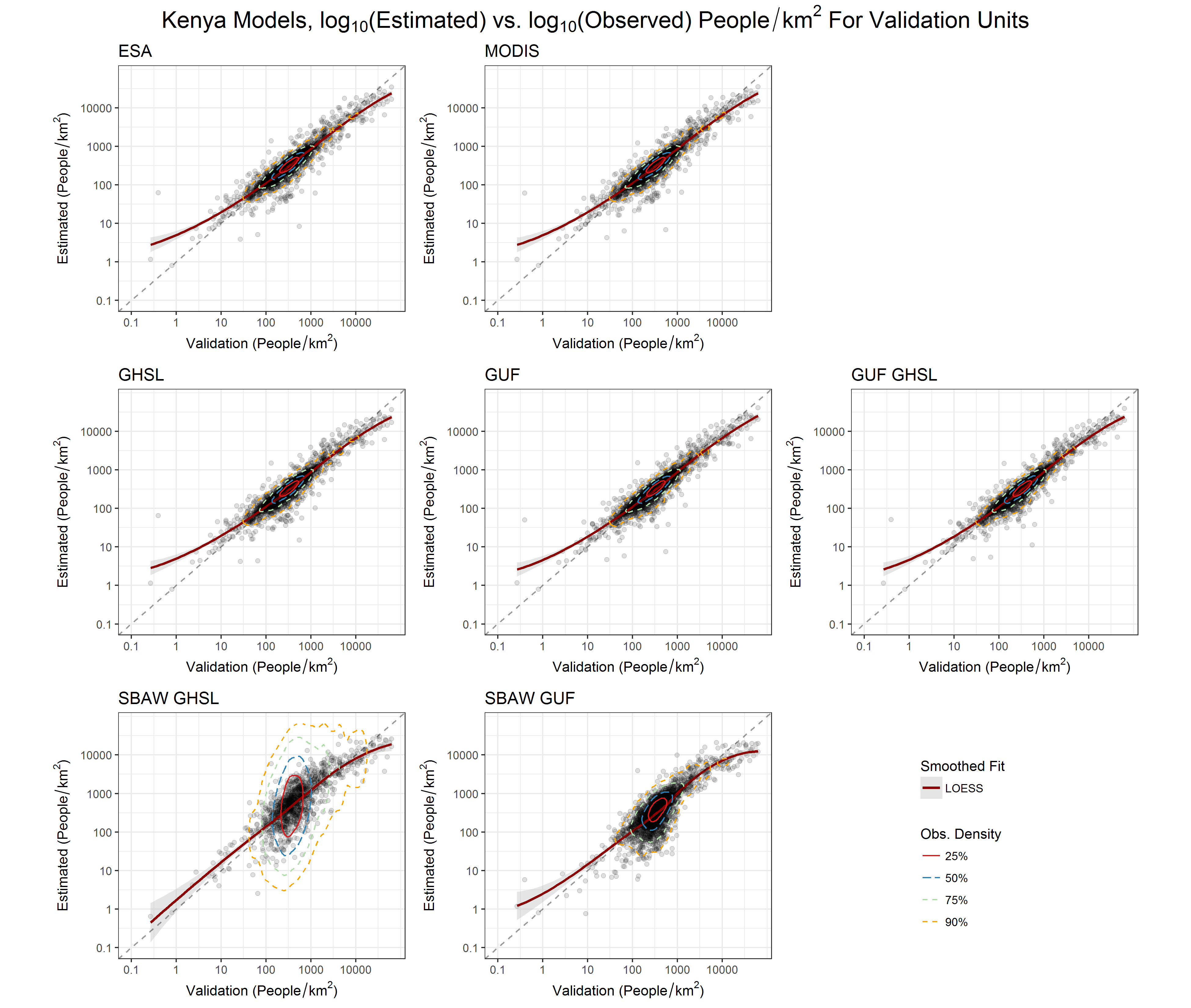


Figure S2.4.

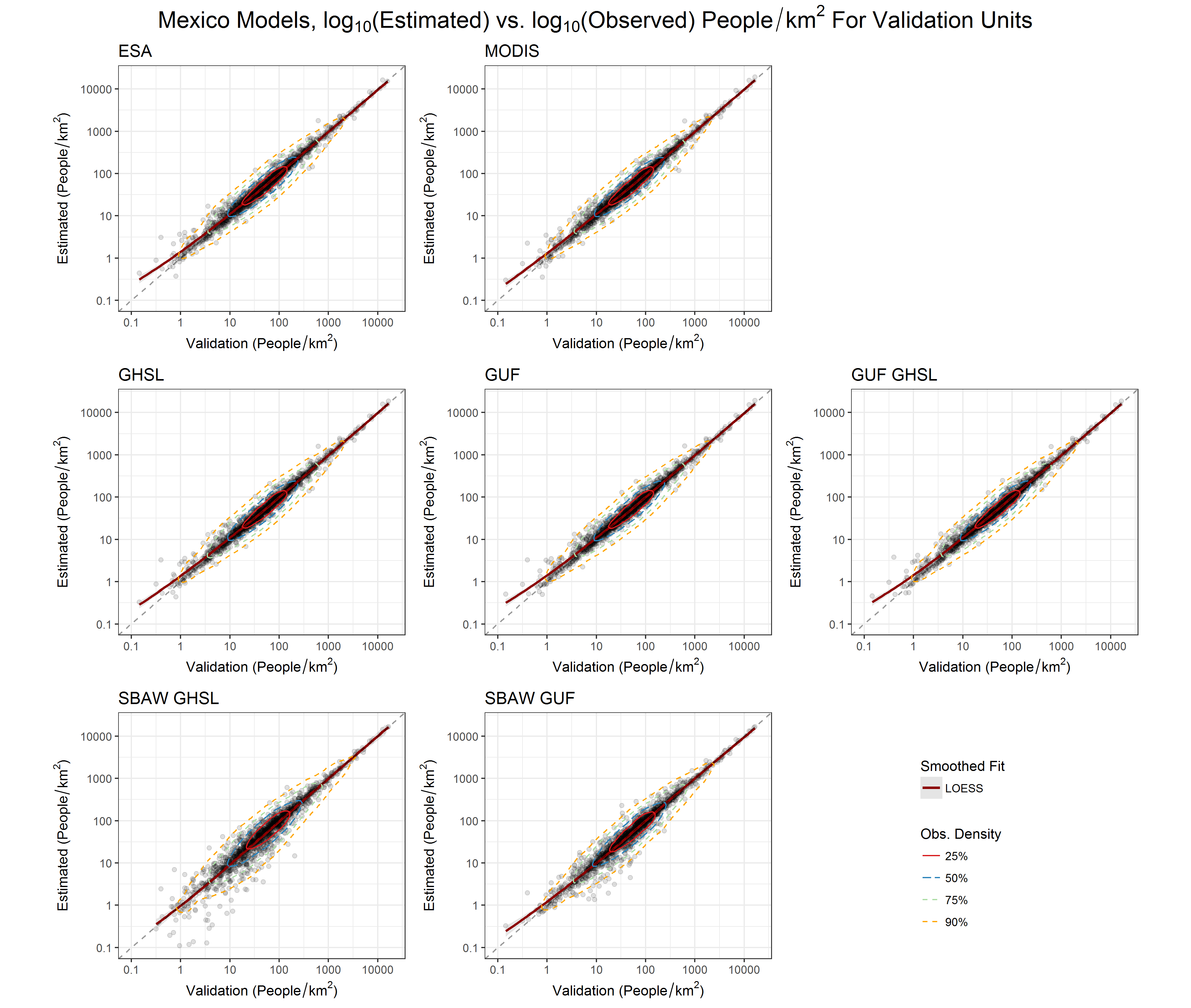


Figure S2.5.

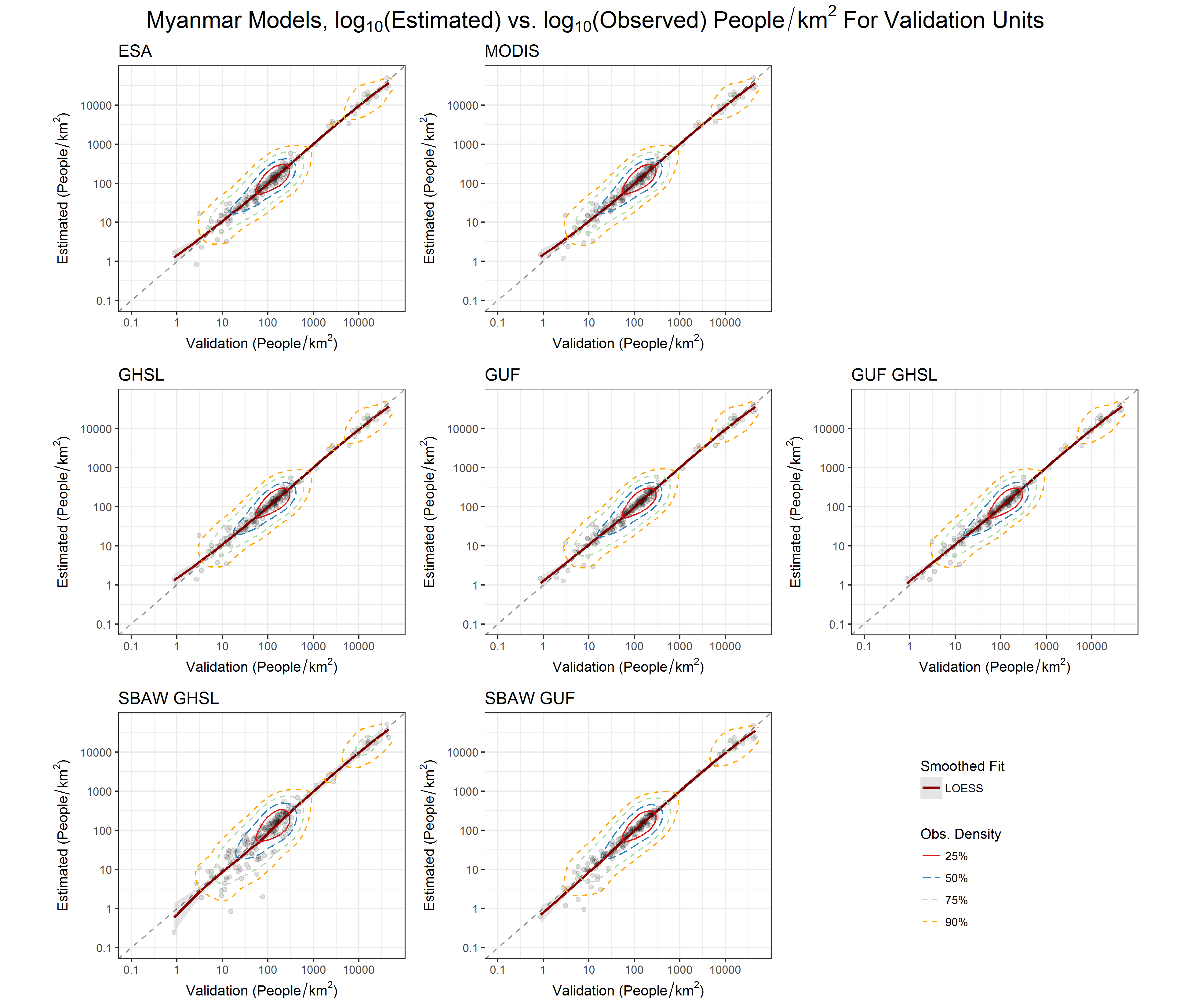


Figure S2.6.

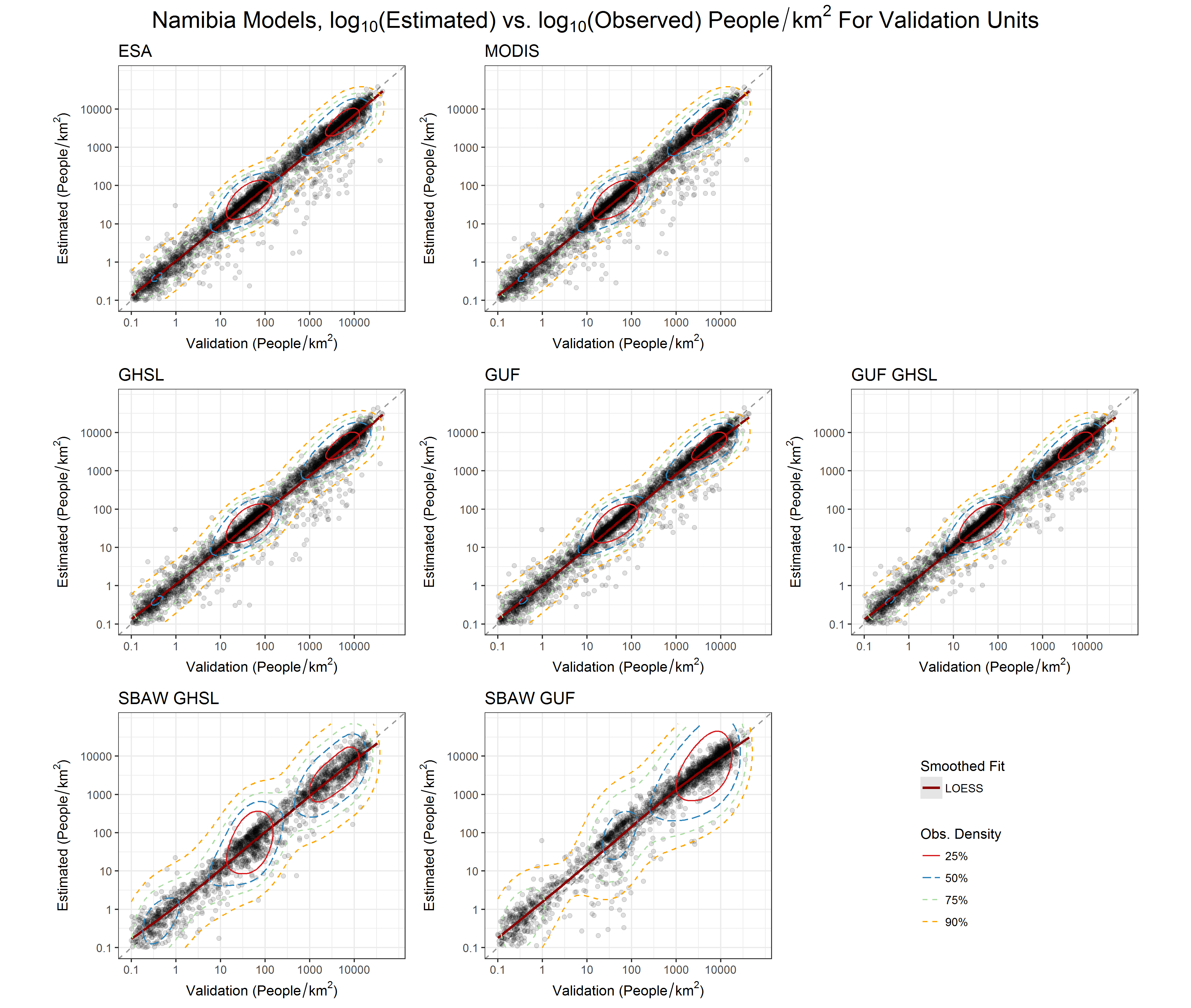


Figure S2.7.

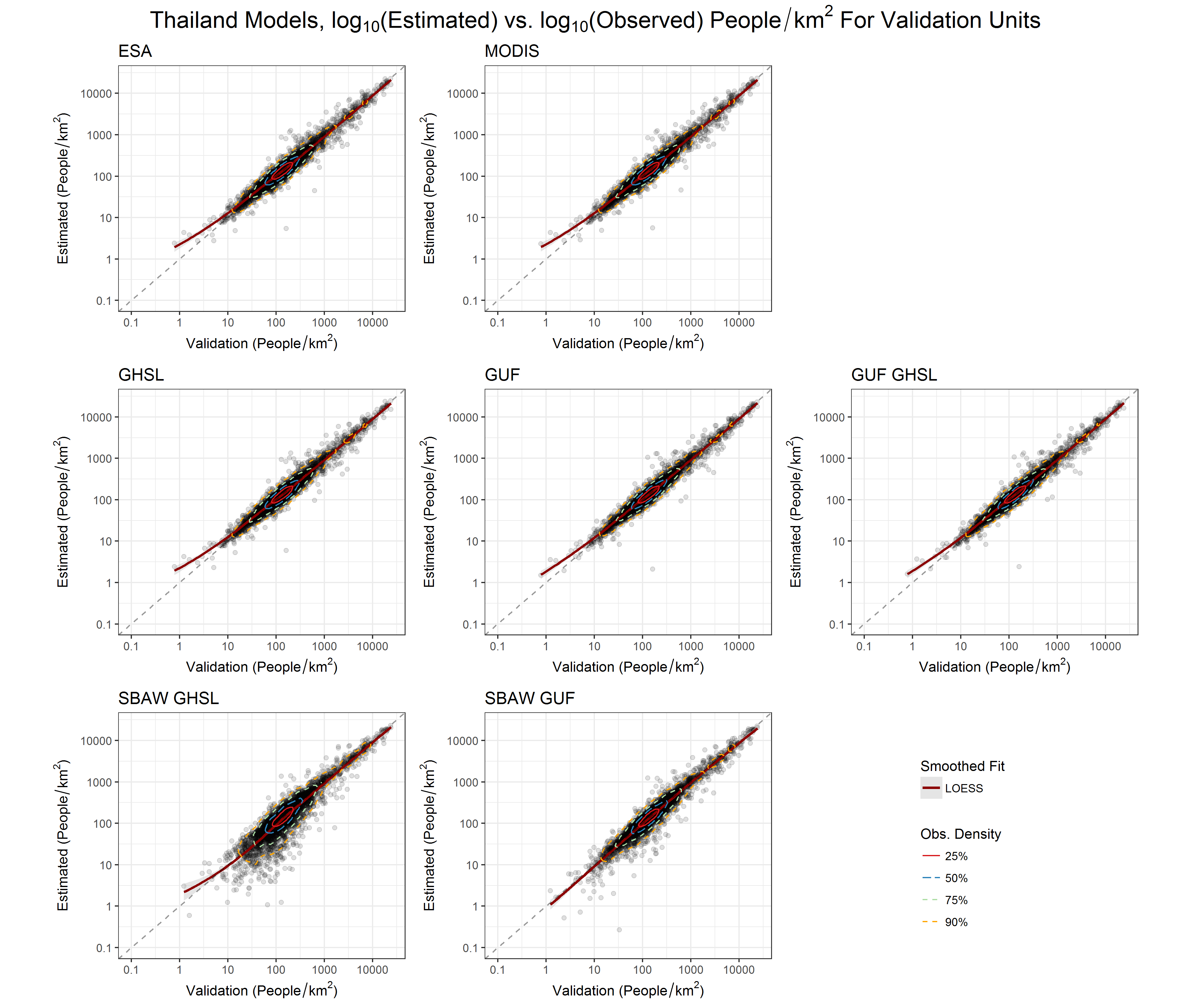


Figure S2.8.

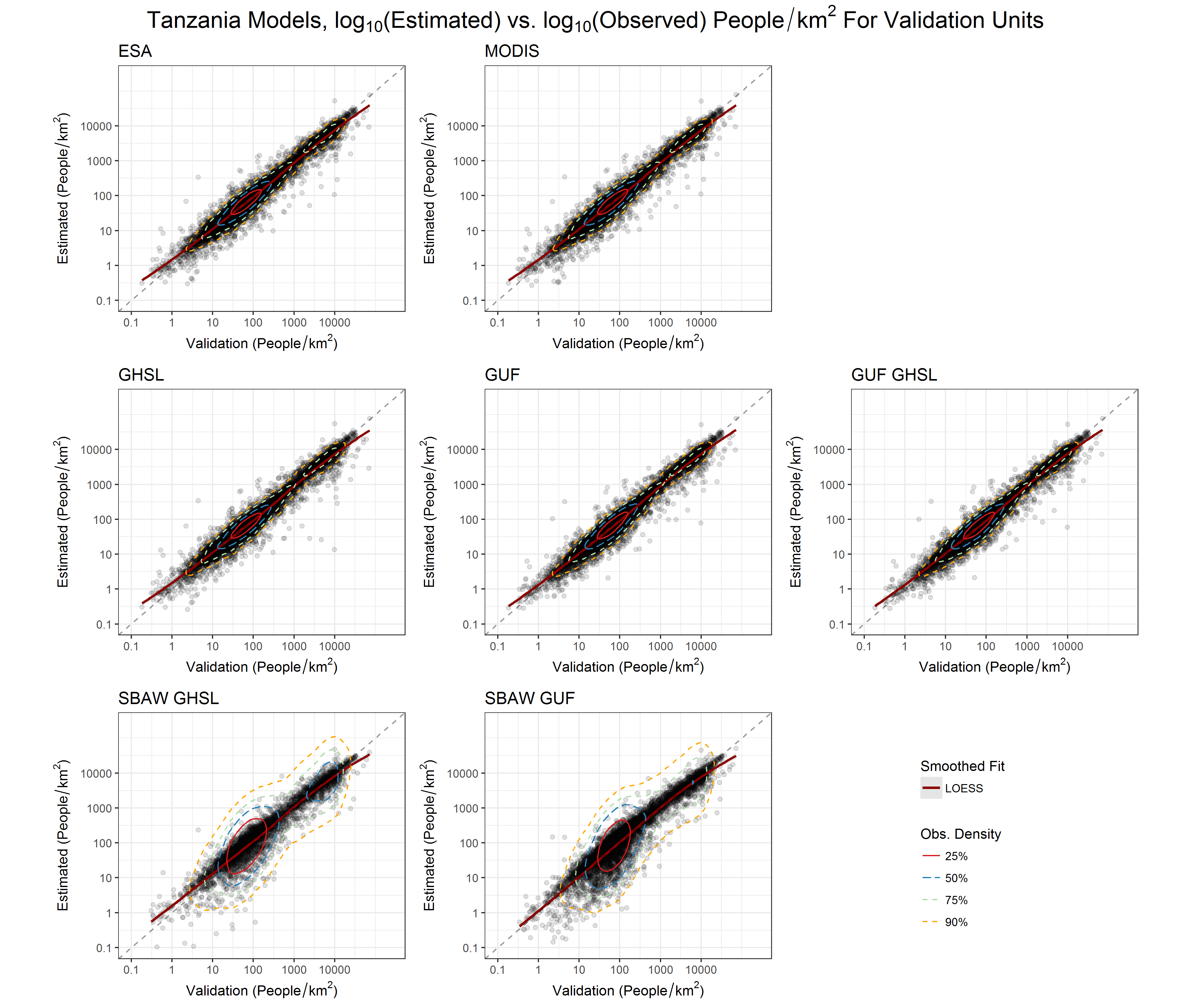


Figure S2.9.

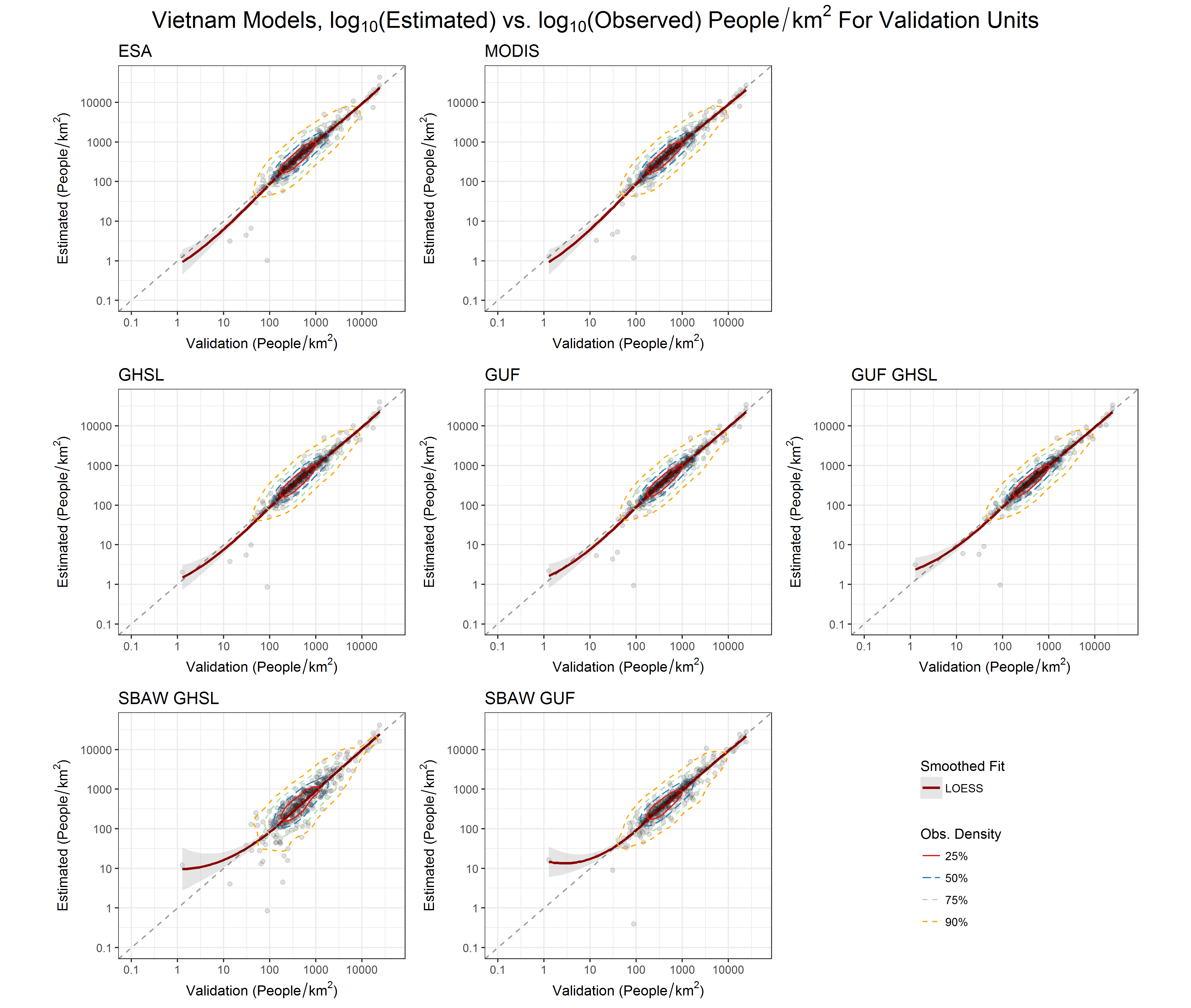


Figure S2.10.

*Supplementary Information S3. Bootstrap Validation Assessment*

We present the bootstrapping results for both root mean square error (RMSE) and mean absolute error (MAE) for all study area countries. The plots are based on “pirate plots,” a boxplot alternative. Along the x-axis are the various model types. The raw data (i.e. 10,000 validation runs) is plotted as points, with their densities distributed along the y-axis. For all figures, the y-axis represents population density in people per hectare.

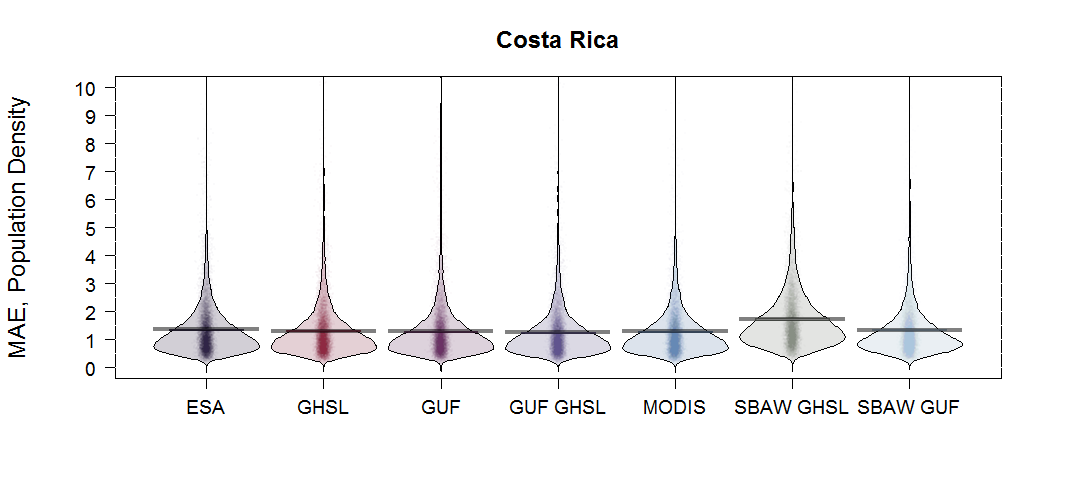
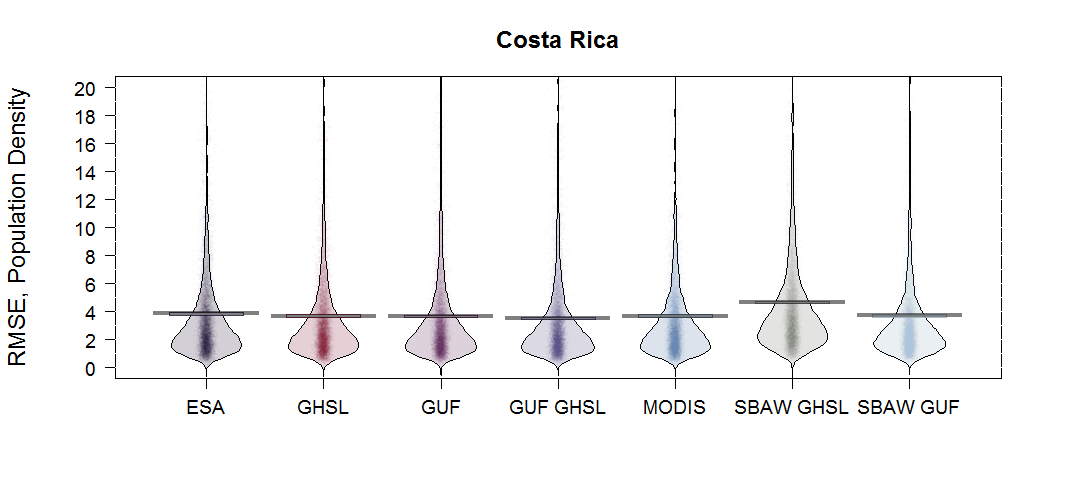


Fig. S3.1. The results by model type for Costa Rica are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.

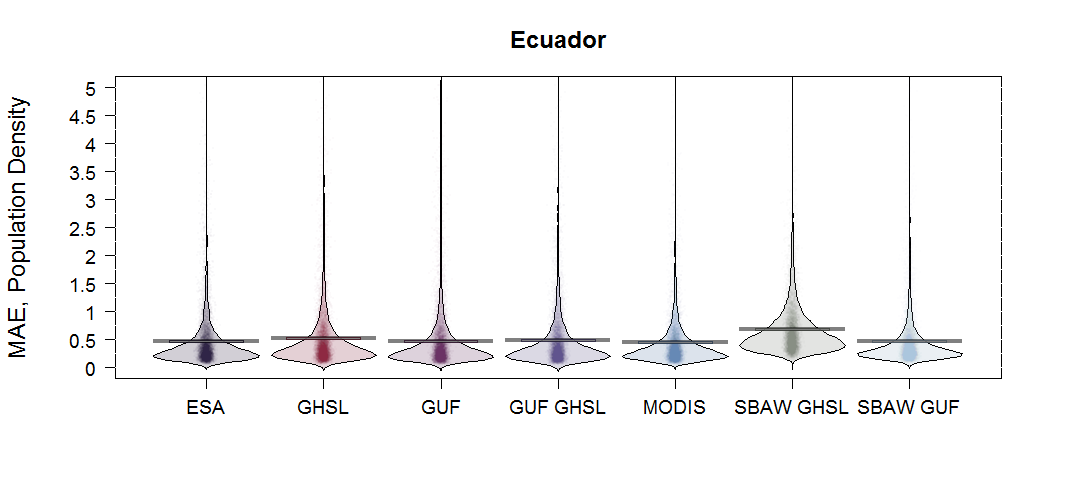
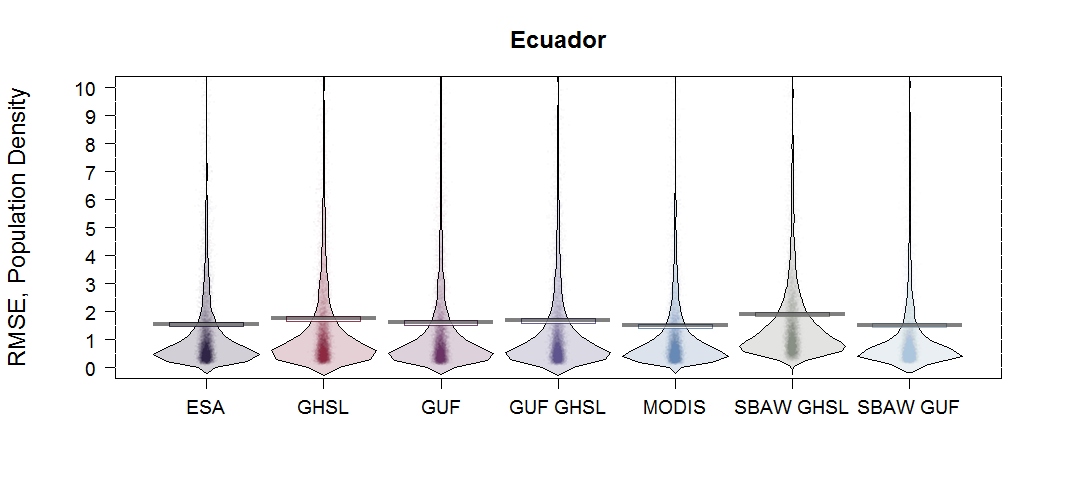


Fig. S3.2. The results by model type for Ecuador are shown for RMSE and MAE metrics. Different models with different built representations are shown on each x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.

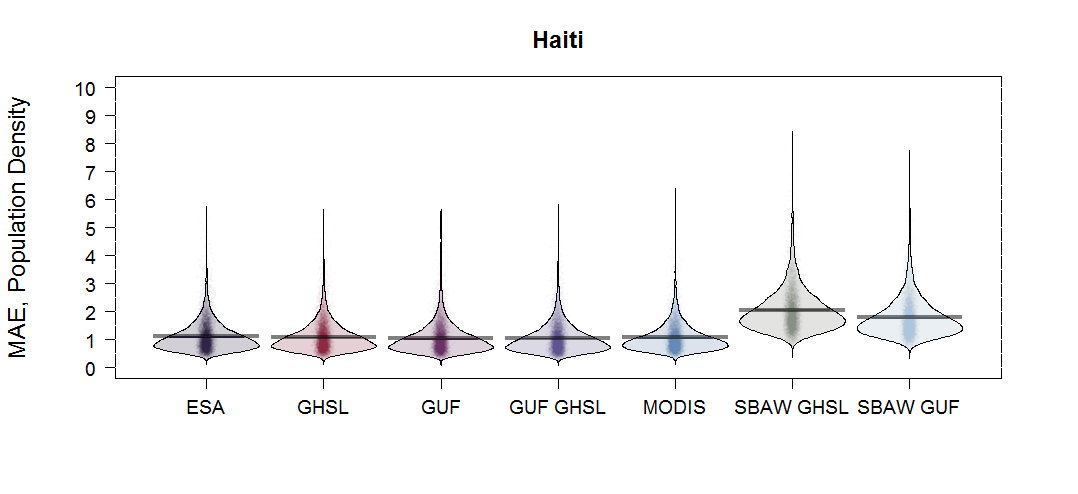
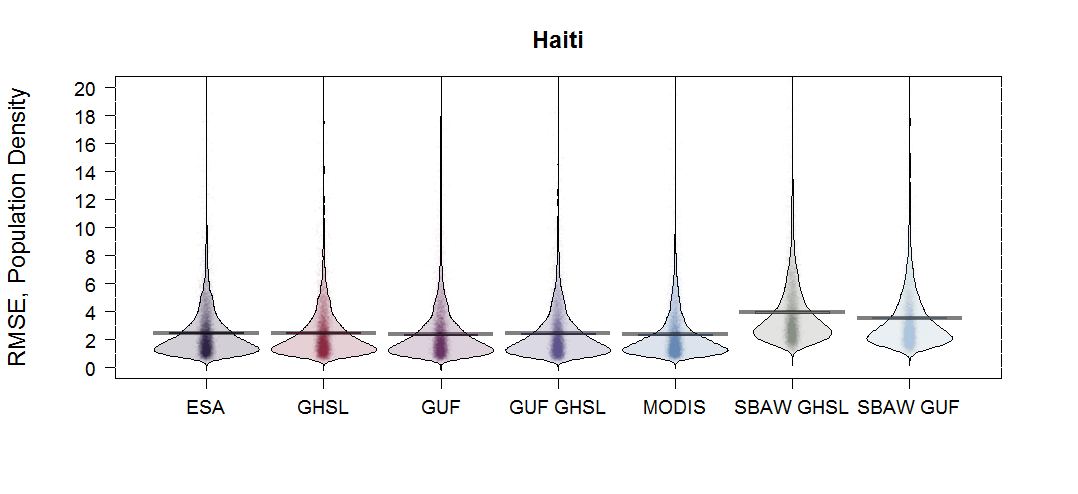
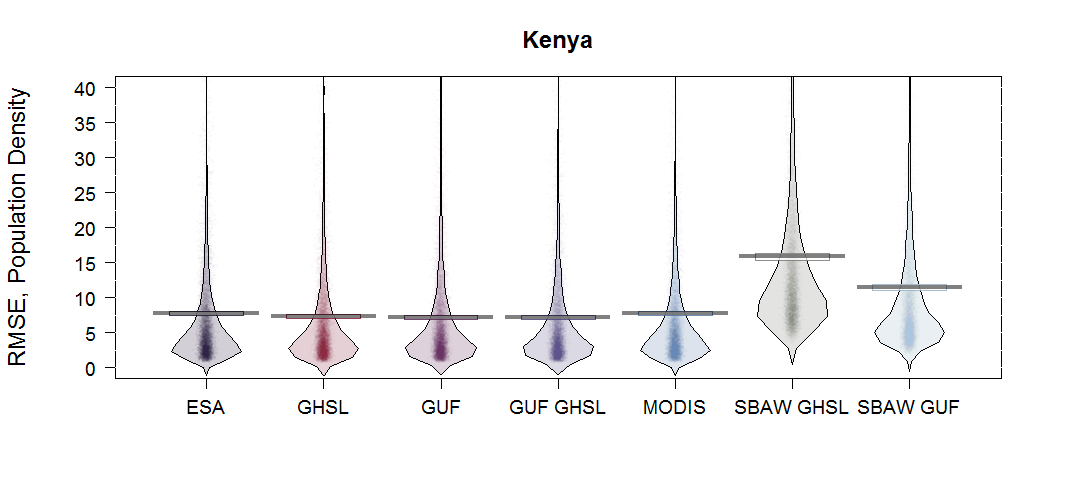


Fig. S3.3. The results by model type for Haiti are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.



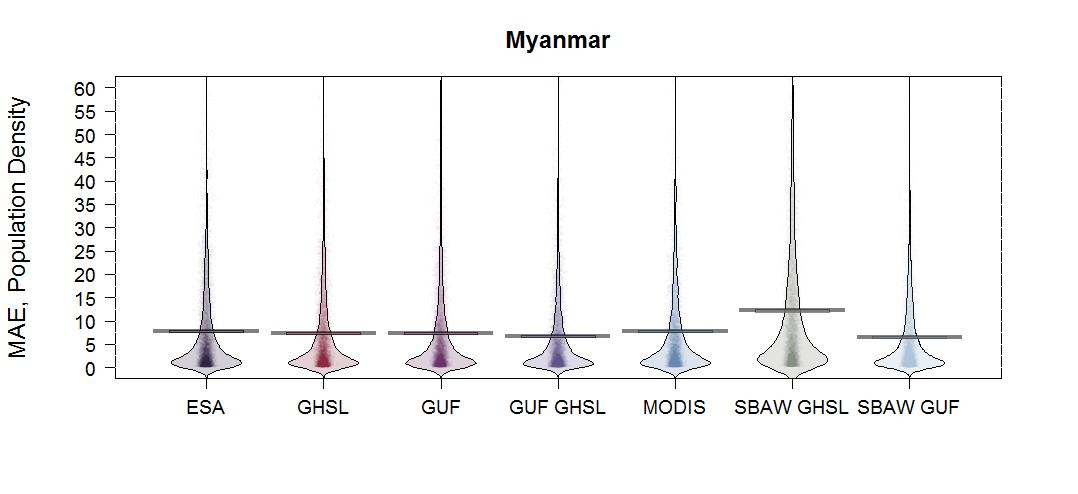


Fig. S3.4. The results by model type for Mexico are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.

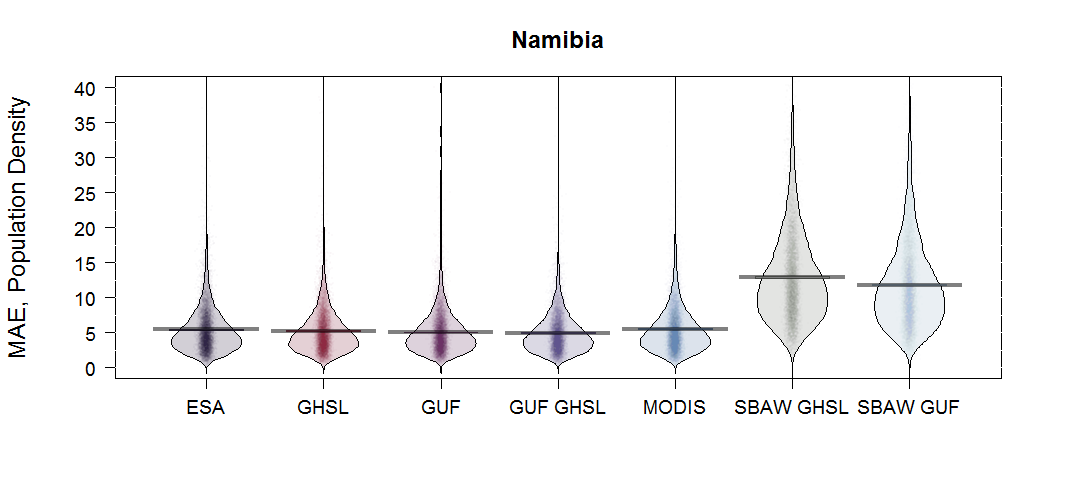
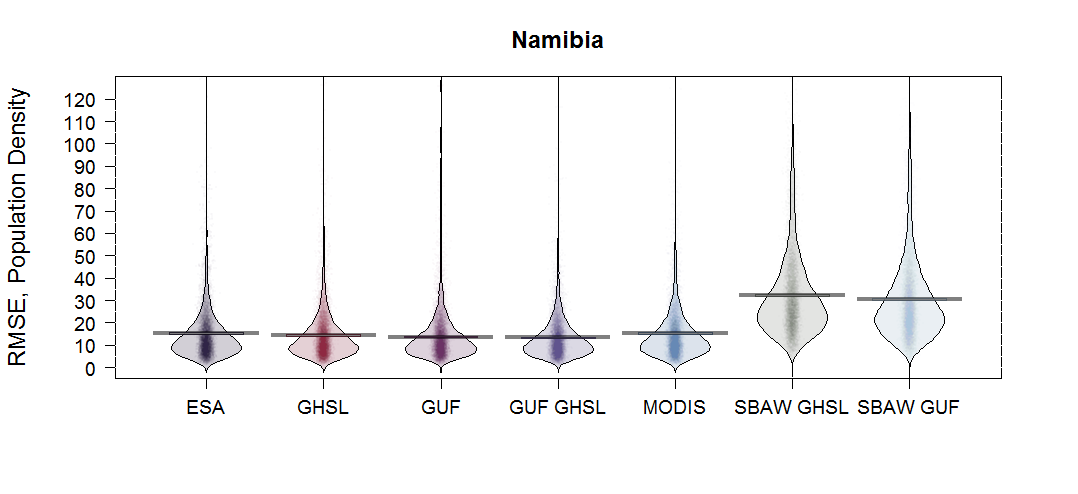


Fig. S3.5. The results by model type for Namibia are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.

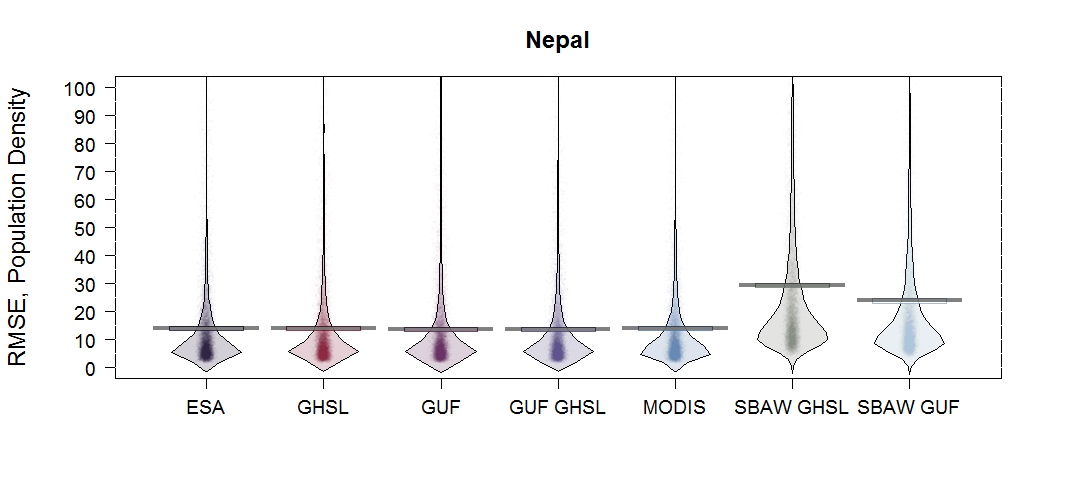
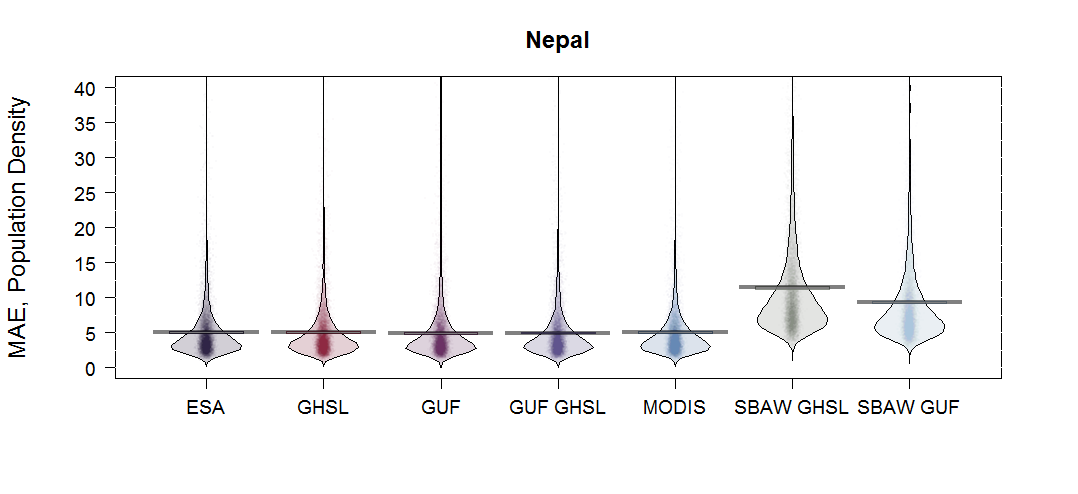
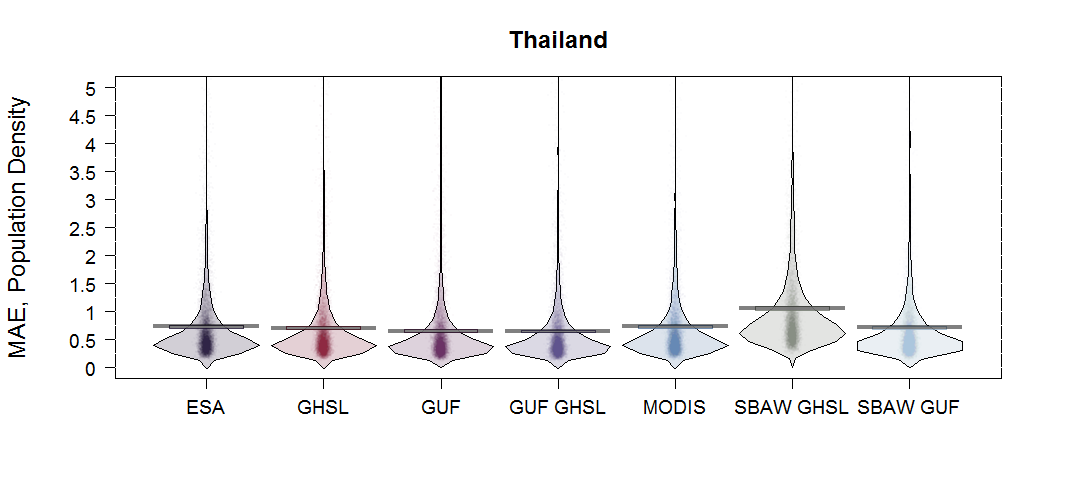


Fig. S3.6. The results by model type for Nepal are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.



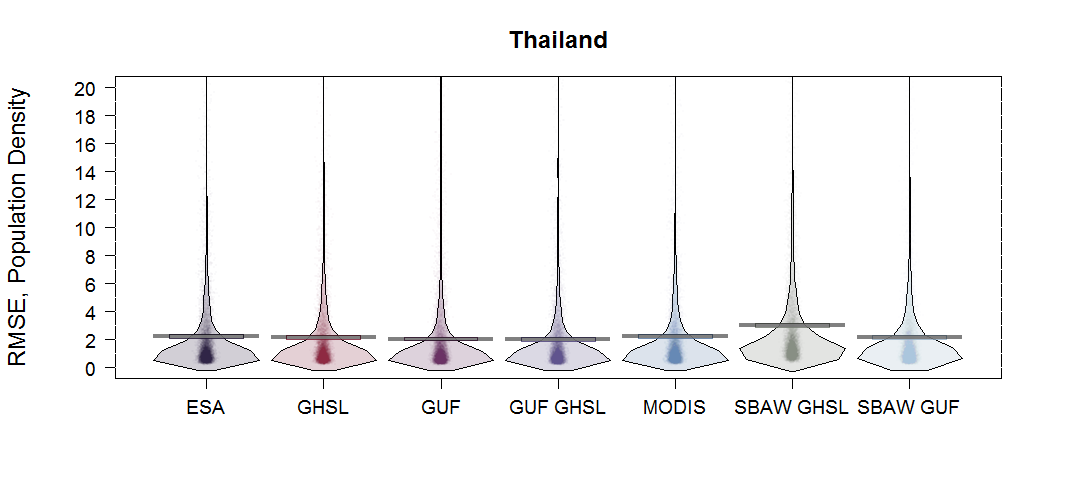


Fig. S3.7. The results by model type for Thailand are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.

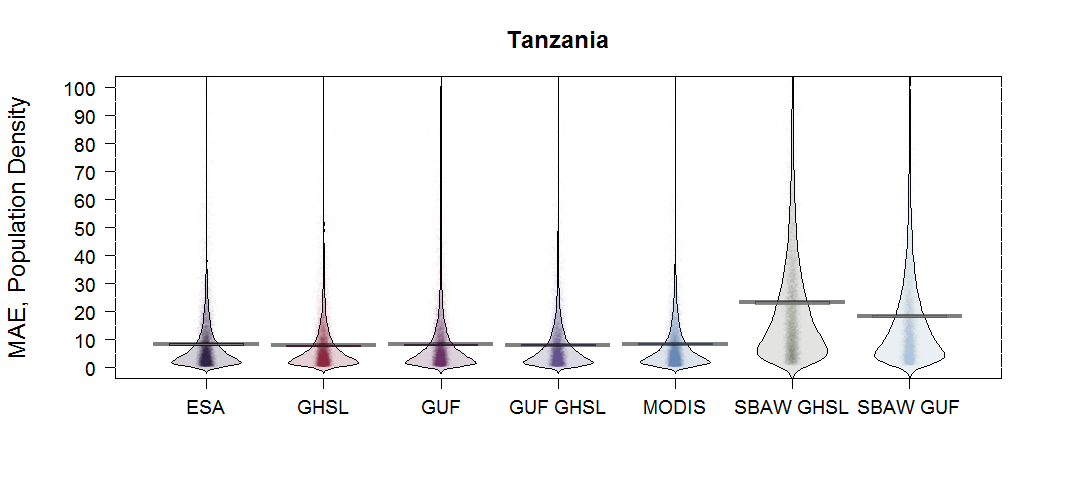
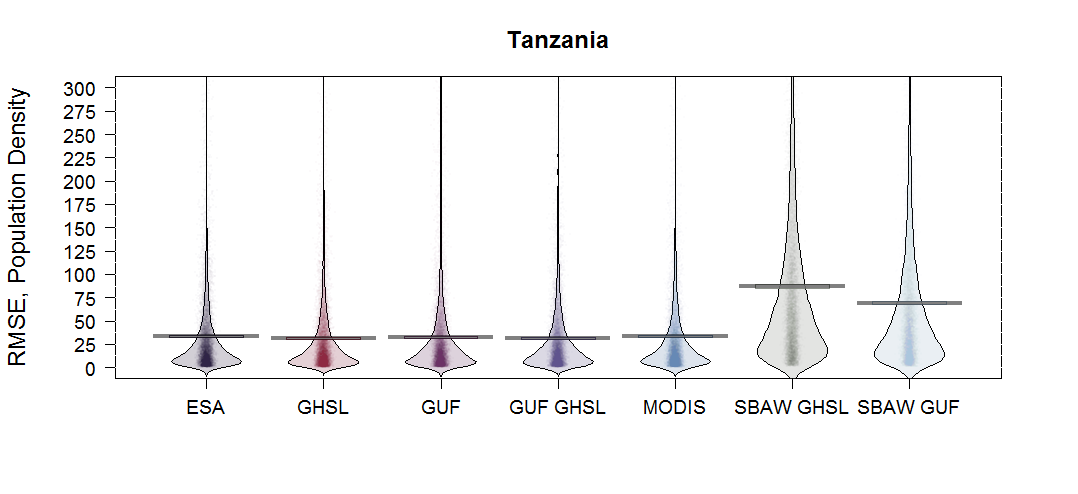
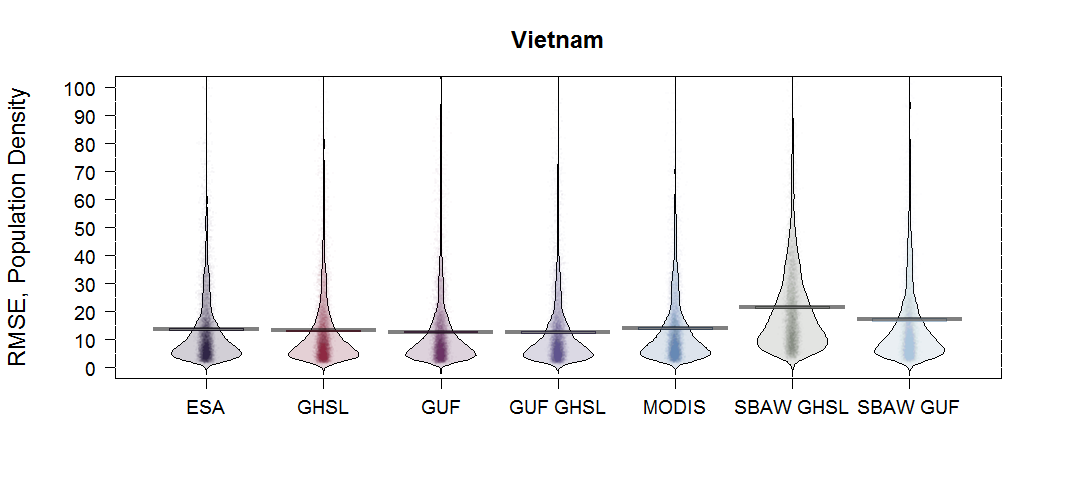


Fig. S3.8. The results by model type for Tanzania are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.



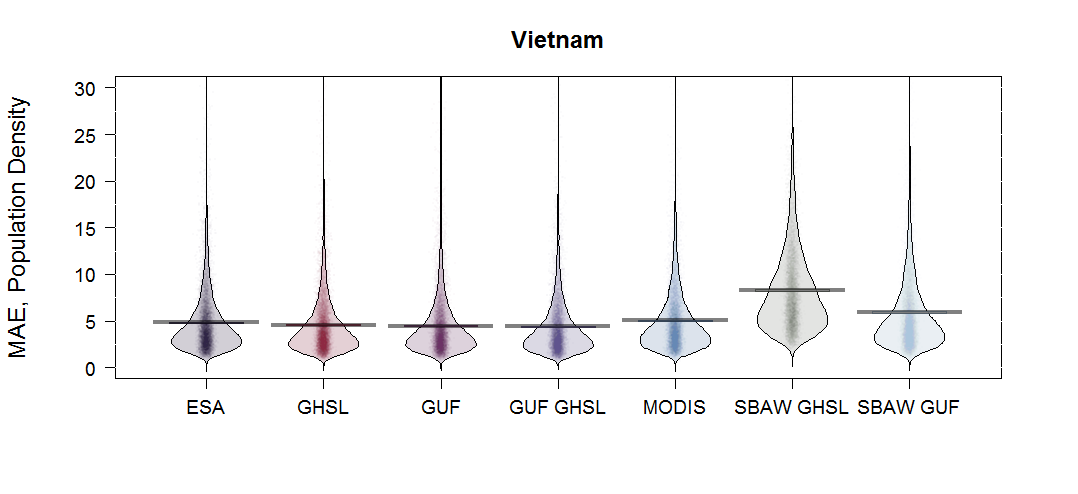


Fig. S3.9. The results by model type for Vietnam are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.

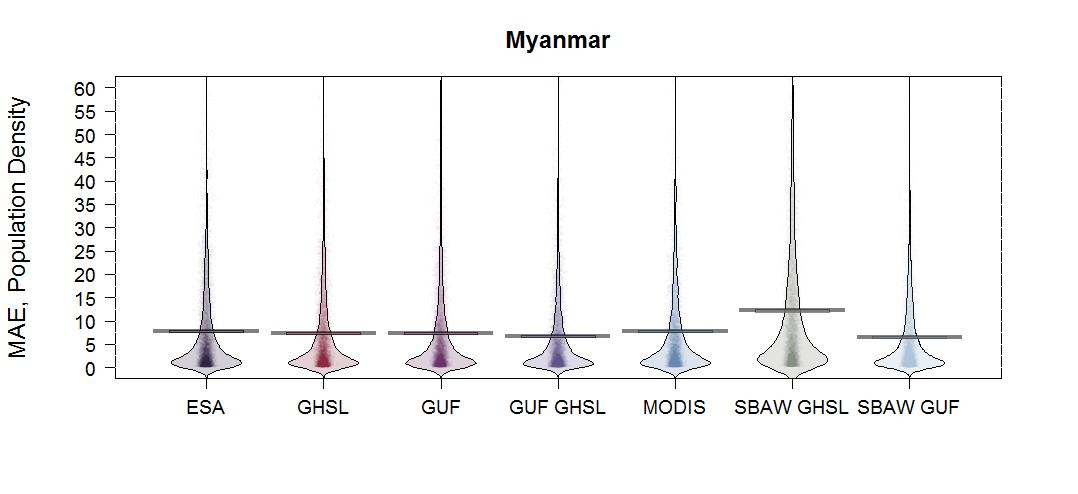
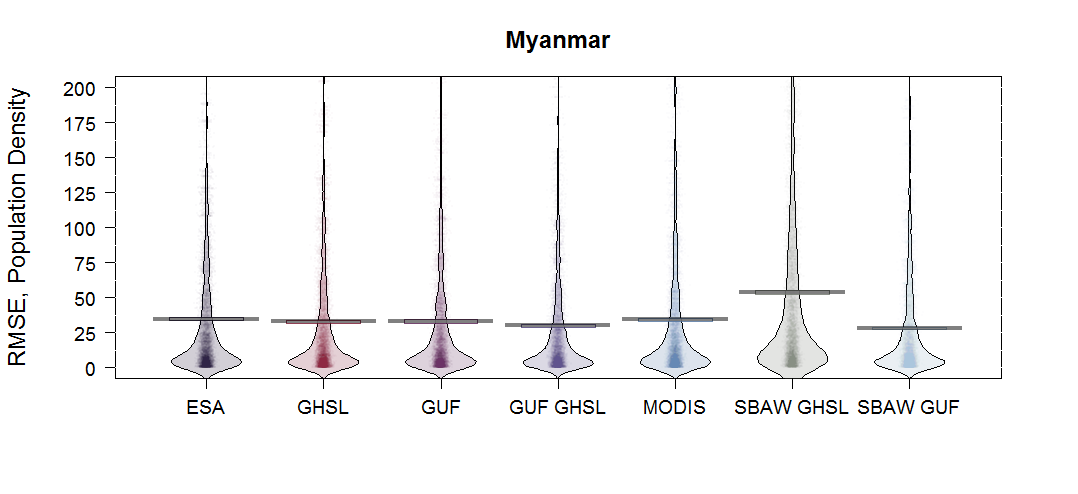


Fig. S3.10. The results by model type for Myanmar are shown for a) RMSE and b) MAE metrics. Different models with different built representations are shown on the x-axis. The colored points represent individual raw data values while the horizontal lines indicate the central tendencies for each model. The first five models are random-forest based while the last two are based on the simple binary areal weighting (SBAW) approach.

*Supplementary Information S4. Census Metadata by Country*

These metadata were compiled for census data used in this study and contributed by the WorldPop Project (<http://www.worldpop.org>). The full metadata reports can be downloaded directly from the website linked above.

**MEXICO Census Data and Observed Population Density**

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

**Mexico Census, 2010**

**Folder:** Census  
**File Name:** CenMex.shp  
**Source:** El Instituto Nacional de Estad�stica y Geograf�a (INEGI), 2010  
**Description:** These high spatial resolution census block data were attained through in-country partners for 2010.  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

nfeatures : 2456

extent : -1107799, 2106369, 1629367, 3702427 (xmin, xmax, ymin, ymax)

coord. ref. : NA

nvariables : 167

## Costa Rica Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Costa Rica Census, 2011

**Folder:** Census  
**File Name:** CRI\_2011\_Census\_Distritos\_Merged\_prj.shp  
**Source:** INEC Instituto Nacional de Estad�stica y Censos de Costa Rica, 2011  
**Description:** These high spatial resolution census block data were attained through combining a downloaded table of Distrito-level census data from INEC and merging them by provincia, canton and distrito with distrito level census data downloaded from ESRI ( <http://www.arcgis.com/home/item.html?id=29462fe665444063b69ac35fa82f4bc0> ). Note that Isla del Coco was removed since it is uninhabitated save for a permanent ranger station, and several edits were made by hand to distrito names and conventions in order to merge successfully with 2011 census counts.  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

nfeatures : 469

extent : 283647, 658874, 889017, 1240761 (xmin, xmax, ymin, ymax)

coord. ref. : NA

nvariables : 17

## Haiti Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Haitian Population Estimate, 2015

**Folder:** Census  
**File Name:** Haiti\_2015\_population\_estimate\_adm3.shp  
**Source:** INSTITUT HA�TIEN DE STATISTIQUE ET D'INFORMATIQUE, 2015  
**Description:** Obtained from HDX, 23-10-2016, <https://data.humdata.org/dataset/estimated-population-of-haiti-2015>  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

Field name: '18PLUS' changed to: 'X18PLUS'

class : SpatialPolygonsDataFrame

features : 570

extent : 554877, 855241, 1992998, 2222861 (xmin, xmax, ymin, ymax)

coord. ref. : NA

variables : 23

## Ecuador Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Ecuador Census, 2010

**Folder:** Census  
**File Name:** ECU\_AdminPop10.shp  
**Source:** Population Census, Instituto National de Estadistica y Censos (INEC), Ecuador, 2010, <http://www.geohive.com/cntry/ecuador.aspx>,  
**Description:** These spatial resolution census block data were attained through in-country partners for 2011.  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

nfeatures : 978

extent : -711683, 1146409, 9445081, 10160826 (xmin, xmax, ymin, ymax)

coord. ref. : NA

nvariables : 22

## Kenya Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Kenya Census Data, 1999, Admin-level 5

**Folder:** Census  
**File Name:** KEN\_census\_1999\_sublocations\_topo.shp  
**Source:** Kenya National Bureau of Statistics, acquired by Tatem, et al. for use in AfriPop data products.  
**Description:** These census data were acquired for use as a disaggregation layer for more-recent census data for AfriPop. It is used here on its own to produce a disaggregated population map for 1999 because it is the finest level census data available. Required fields for map production are ADMINID and ADMINPOP.  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

nfeatures : 6624

extent : -66764, 823501, -517009, 605783 (xmin, xmax, ymin, ymax)

coord. ref. : NA

nvariables : 50

## Namibia Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Namibian Census Data, 2011, Admin-level 2

**Folder:** Census  
**File Name:** adminpop2011.shp  
**Source:** Namibia Statistics Agency, the Republic of Namibia  
**Description:** These census data were acquired in April 2014 for use in WorldPop endeavors. Required fields for map production are ADMINID and ADMINPOP  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

nfeatures : 107

extent : -165880, 1271387, -3205127, -1881886 (xmin, xmax, ymin, ymax)

coord. ref. : NA

nvariables : 19

## Nepal Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Nepal Census Data, 2011, Admin-level 5

**Folder:** Census  
**File Name:** census.shp  
**Source:** Central Bureau of Statistics of Nepal  
**Description:** These census data were acquired in April 2014 for use in WorldPop endeavors. Required fields for map production are ADMINID and ADMINPOP  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

features : 36036

extent : 115788, 916842, 2918485, 3370757 (xmin, xmax, ymin, ymax)

coord. ref. : NA

variables : 18

## Myanmar Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Myanmar Census Data, 2014, Admin-level 3

**Folder:** Census  
**File Name:** MMR\_2014\_Census\_Admin\_lvl\_3\_hybrid.shp  
**Source:** Republic of the Union of Myanmar Ministry of Immigration and Population  
**Description:** These census data were acquired in July 2015 for use in WorldPop endeavors. To correct for population counts the Sittwe, Maungdaw, and Mrauk-U districts of Rakhine state were merged and therfore their input census resolution lay between admin lvl 1 and 2. Further, the townships of Momauk and Mansi were merged into the Bhamo district, of Kachin state, and had an input census resolution of admin lvl 2. Lastly, because of incompatibilities between the shapefile admin boundary data and the township level census data, areas corresponding to the Wa self-administered regions in the Shan state were merged into districts and input into the model at admin lvl 2. All other areas were input into the model as admin lvl 3 (township) data.  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

features : 316

extent : 102470, 1035661, 1069650, 3158647 (xmin, xmax, ymin, ymax)

coord. ref. : NA

variables : 10

## Vietnam Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Vietnam Census Data, 1999, Admin-level 4

**Folder:** Census  
**File Name:** VNM\_popdata\_final.shp  
**Source:** General Statistics Office, Vietnam, acquired by Gaughan, et al. for use in AsiaPop data products.  
**Description:** These census data were acquired for AsiaPop (Gaughan, et al., 2013). Required fields for map production are ADMINID and ADMINPOP.  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

nfeatures : 10613

extent : 206025, 985767, 945724, 2587027 (xmin, xmax, ymin, ymax)

coord. ref. : NA

nvariables : 30

## Thailand Census Data and Observed Population Density

These data are the population density values used to estimate the RF model used to create the prediction weighting layer you see above. Values represent population density as measured by people per hectare and calculated from population counts within each census unit. These values are used as the dependent variable during model estimation.

### Thailand Census Data, 2010, Admin-level 3

**Folder:** Census  
**File Name:** Hupop\_SubdistrictThailand\_2010\_wgs84.shp  
**Source:** National Statistics Office of Thailand, acquired by Gaughan, et al. for use in AsiaPop data products.  
**Description:** These census data were acquired for AsiaPop (Gaughan, et al., 2013). Required fields for map production are ADMINID and ADMINPOP.  
**Class:** polygon  
**Derived Covariates:**  
area, buff, zones,

class : SpatialPolygonsDataFrame

nfeatures : 7416

extent : -314299, 568270, 621843, 2271005 (xmin, xmax, ymin, ymax)

coord. ref. : NA

nvariables : 12