

Article

Using Information on Settlement Patterns to Improve the Spatial Distribution of Population in Coastal Impact Assessments

Jan-Ludolf Merkens *  and Athanasios T. Vafeidis

Department of Geography, Kiel University, Ludewig-Meyn-Str. 14, 24118 Kiel, Germany; vafeidis@geographie.uni-kiel.de

* Correspondence: merkens@geographie.uni-kiel.de; Tel.: +49-431-880-1701

Received: 15 July 2018; Accepted: 3 September 2018; Published: 5 September 2018



Abstract: Broad-scale impact and vulnerability assessments are essential for informing decisions on long-term adaptation planning at the national, regional, or global level. These assessments rely on population data for quantifying exposure to different types of hazards. Existing population datasets covering the entire globe at resolutions of 2.5 degrees to 30 arc-seconds are based on information available at administrative-unit level and implicitly assume uniform population densities within these units. This assumption can lead to errors in impact assessments and particularly in coastal areas that are densely populated. This study proposes and compares simple approaches to regionalize population within administrative units in the German Baltic Sea region using solely information on urban extent from the Global Urban Footprint (GUF). Our results show that approaches using GUF can reduce the error in predicting population totals of municipalities by factor 2 to 3. When assessing exposed population, we find that the assumption of uniform population densities leads to an overestimation of 120% to 140%. Using GUF to regionalise population within administrative units reduce these errors by up to 50%. Our results suggest that the proposed simple modeling approaches can result in significantly improved distribution of population within administrative units and substantially improve the results of exposure analyses.

Keywords: spatial population; Global Urban Footprint; Dasymetric Mapping; coastal exposure; impact assessment; Baltic Sea

1. Introduction

Coastal areas are highly exposed to natural hazards [1] and this exposure will increase as a result of climate-change-induced sea-level rise (SLR) and associated impacts such as flooding, erosion, permanent inundation, and saltwater intrusion [2]. Coastal flooding, in particular, will increase in frequency and intensity [3] and is expected to be the most costly impact of SLR [4]. At the same time, rapid socioeconomic development [5–7], leading to high concentration of people assets in coastal regions and particularly in large urban centers, is expected to further exacerbate flood risk.

Adaptation measures, in the form of protection, accommodation or retreat, can reduce the impact of coastal flooding by several orders of magnitude [4]. In this context, establishing long-term adaptation policies constitutes a key element for the sustainability of coastal regions [8] and will be necessary for achieving the United Nations Sustainable Development Goals (SDG) outlined in the “2030 Agenda” [9]. In particular, coastal adaptation explicitly relates to SDGs 11 (Sustainable Cities and Communities), 13 (Climate Action), and 14 (Life below Water) and is indirectly linked to other SDGs.

To inform decisions on coastal adaptation, assessment of the impact of coastal flooding has been carried out at different scales [4,10–12], from global to local. An essential input, but also one

of the main sources of uncertainty for these assessments, is the spatial distribution of population. This parameter is important for defining exposure to hazards, particularly in global and regional studies. However, available global datasets are limited in the way they represent how people are distributed in space. Two of the most commonly employed datasets are the Gridded Population of the World (GPW) [13] and the Global Rural Urban Mapping Project (GRUMP) [14]. They cover the entire globe at a spatial resolution of 30 arc-seconds (approximately 1 km at the equator) and have been widely used in coastal-exposure analysis (e.g., References [7,15,16] for GPW; References [5,6,17] for GRUMP). GPW is based on population census tables of administrative units and distributes population uniformly on land areas within administrative units [18]. GRUMP additionally differentiates between rural and urban areas, which are derived from nightlight satellite images. This method performs better in developed regions than in developing or undeveloped regions due to lack of electricity and less light pollution [19]. The assumption of a uniformly distributed population is, however, rather crude and does not necessarily represent the true distribution of population. Dasymetric-mapping approaches aim to overcome this limitation by using ancillary data to spatially differentiate population within administrative units [20]. The algorithms used in dasymetric mapping can become very complex and data-demanding. WorldPop, for example, which provides freely available gridded population data for Latin America [21], Asia [22], and Africa [23], uses about 40 variables to model population densities [24]. The spatial resolution of these datasets is 30 arc-seconds for data on the continental level and 3 arc-seconds (approximately 100 m at the equator) for country-level data. For global or continental studies, finer spatial resolution would be desirable, as the use of data of higher spatial resolution can result in more detailed impact assessments [25]. However, collecting data for such large numbers of variables is time- and resource-intensive, and very difficult to implement on global scales.

In this study, we used simple modeling approaches that were based on low data requirements, for producing improved estimates of population distribution at high spatial resolution. We applied and compared these approaches in the German Baltic Sea coast by assessing exposure of population to coastal flooding. To ensure transferability to other areas of interest, we used solely the Global Urban Footprint (GUF) [26,27] as ancillary data for regionalizing population within census units. The GUF is a binary settlement mask that covers the entire globe. It is available at spatial resolutions of 2.8 arc-seconds (~84 m at the equator) and 0.4 arc-seconds (~12 m at the equator).

We further investigated which of the two GUF produce result in more realistic spatial population patterns, thus providing improved estimates of population exposure to coastal flooding. For evaluating the proposed approaches, we compared the estimates of exposed population produced with the use of modeled population distributions to the actual exposed population based on gridded census population data. Although the focus of this study is on coastal impact assessment, reliable estimates of the spatial distribution of population have a wide range of applications, such as impact assessments of other natural hazards or quantification of exposure of population to diseases.

2. Study Area

We analyzed 17 districts (NUTS-3 level) in the federal states of Schleswig-Holstein and Mecklenburg-Vorpommern in northern Germany that extend to the Baltic Sea. Ten of the districts are rural (see Figure 1) and consist of 816 municipalities (LAU-2 level). The seven urban districts consist of one municipality each, which leads to 823 municipalities in total. Circa 2.25 million people live in the study area, which encompasses approximately 17,000 km². With a mean population density of about 130 people per km², the study area is predominantly rural. The three cities with the highest population are Kiel (235,000), Lübeck (210,000), and Rostock (200,000) [28]. Including islands, the coastline has a length of approximately 2600 km [29]. About a quarter of the coast is protected by dikes, which mostly protect densely populated areas [30]. Storm surges constitute the primary coastal hazard in the region, with the highest storm surge recorded occurring in 1872 with a measuring a peak height of 3.30 m in Travemünde (Schleswig-Holstein) [31] and 2.83 m in Wismar (Mecklenburg-Vorpommern) [29].

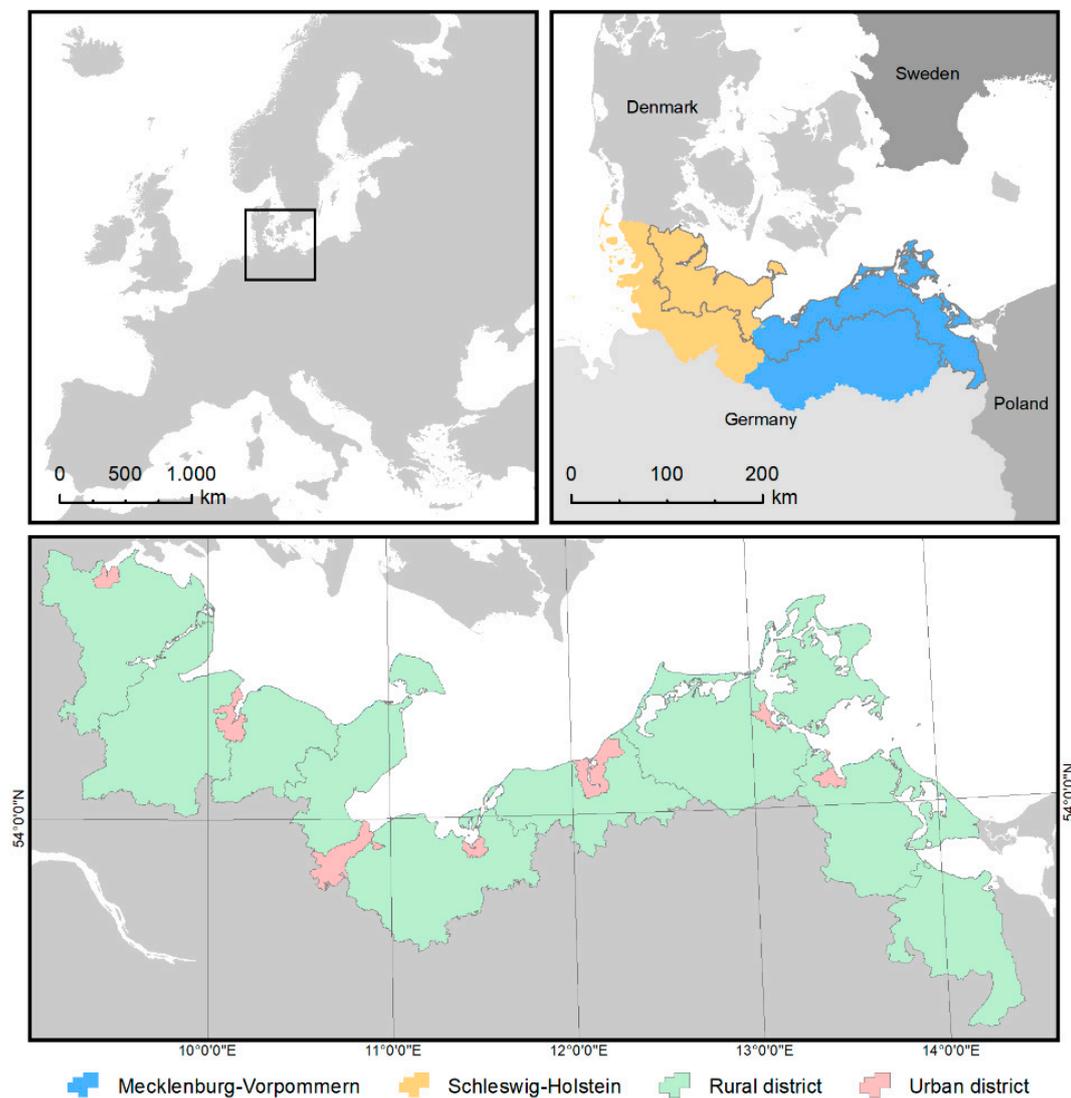


Figure 1. Location, administrative boundaries, and distribution of urban/rural districts for the study area.

3. Data and Methods.

3.1. Data

3.1.1. Population

We used two products of the 2011 Census. First, we used the regular population grid with a horizontal resolution of 100 m [32]. The population density (population per hectare) and population count (population per cell) did not differ due to the Lambert Azimuthal Equal Area Projection (LAEA-EPSC:3035) used in this study. For the 100 m population grid, all cells with population below 3 people per hectare were adjusted for data-protection reasons; namely, the values of cells with a population of 2 have been adjusted to 3, and cells with a population of 1 have been adjusted to 0 [33,34]. To assess the effects of this data manipulation on the total population numbers in the study area, we aggregated the population in the grid of the entire study area. We compared the totals to the population count reported on the municipality level [28], which is the finest level of census data and the second product of the 2011 Census used in this study. The sum of the gridded population exceeded the sum of population reported per administrative unit by 6404 for the entire study area.

As this was a share of circa 0.3% on the total population, we considered the 100 m population grid as reliable reference data for evaluating different population regionalization approaches.

3.1.2. Urban Extent

For the identification of urban areas, we employed the GUF, which is a binary settlement mask that has been derived from TanDEM-X and TerraSAR-X radar images collected in 2011 and 2012 [27]. It is a global dataset, which is available in two resolutions, namely 2.8 arc-seconds (hereinafter referred to as GUF2.8), and 0.4 arc-seconds (hereinafter referred to as GUF0.4). At the equator, this corresponds to a resolution of approximately 84 m and 12 m, respectively.

We projected both GUF2.8 and GUF0.4 to LAEA (see Figure 2). For GUF2.8 we used the cell positions of the census population raster and assigned the value of the nearest neighbor. For GUF0.4, we also used the cell positions of the census population raster and calculated the share of urban GUF0.4 cells assigned to each cell with a spatial resolution of 100 m. For the original GUF0.4, we used all cells with an urban share larger than 0% and classified them as urban. Additionally, we created an urban mask (hereinafter referred to as GUF0.4_{5%}) that employed a threshold of 5% to classify cells as urban (see Section 4 for discussion on the threshold).

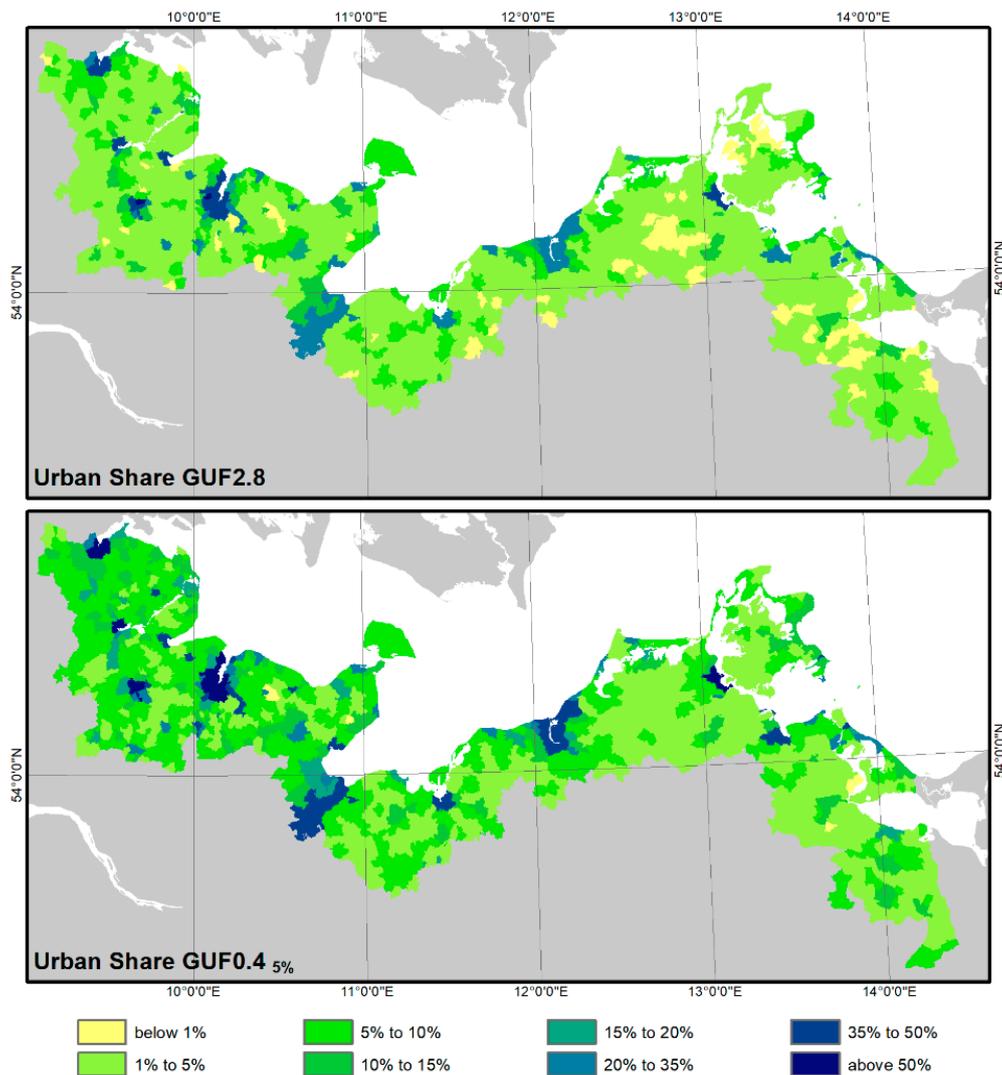


Figure 2. Urban share per municipality for Global Urban Footprint (GUF)2.8 and GUF0.4_{5%}. We defined urban share as the percentage of land area classified as urban.

3.1.3. Exposed Area

We used a Digital Elevation Model (DEM) representing surface heights with a vertical accuracy of approximately 15 cm and a horizontal resolution of 1 m [35] to assess the exposed area (see Figure 3). In a first step, we projected the data to the LAEA projection and aggregated the data to a spatial resolution of 10 m to reduce processing time. Next, we estimated the coastal floodplain considering 8-sided hydrological connectivity to the sea [36]. We used a threshold of 3 m, which corresponds approximately to the height of the 1872 storm surge. As all non-river-induced floodplains were within a distance of 24 km to the coastline, we additionally implemented a maximum flow distance of 24 km from the Baltic Sea to prevent overestimation of the floodplain due to rivers and channels. In a last step, we further aggregated the data to 100 m to match the spatial resolution of the population data. In the aggregation process, we calculated the share of cells representing exposure per hectare and assessed exposed population by multiplying the calculated share with the population count.

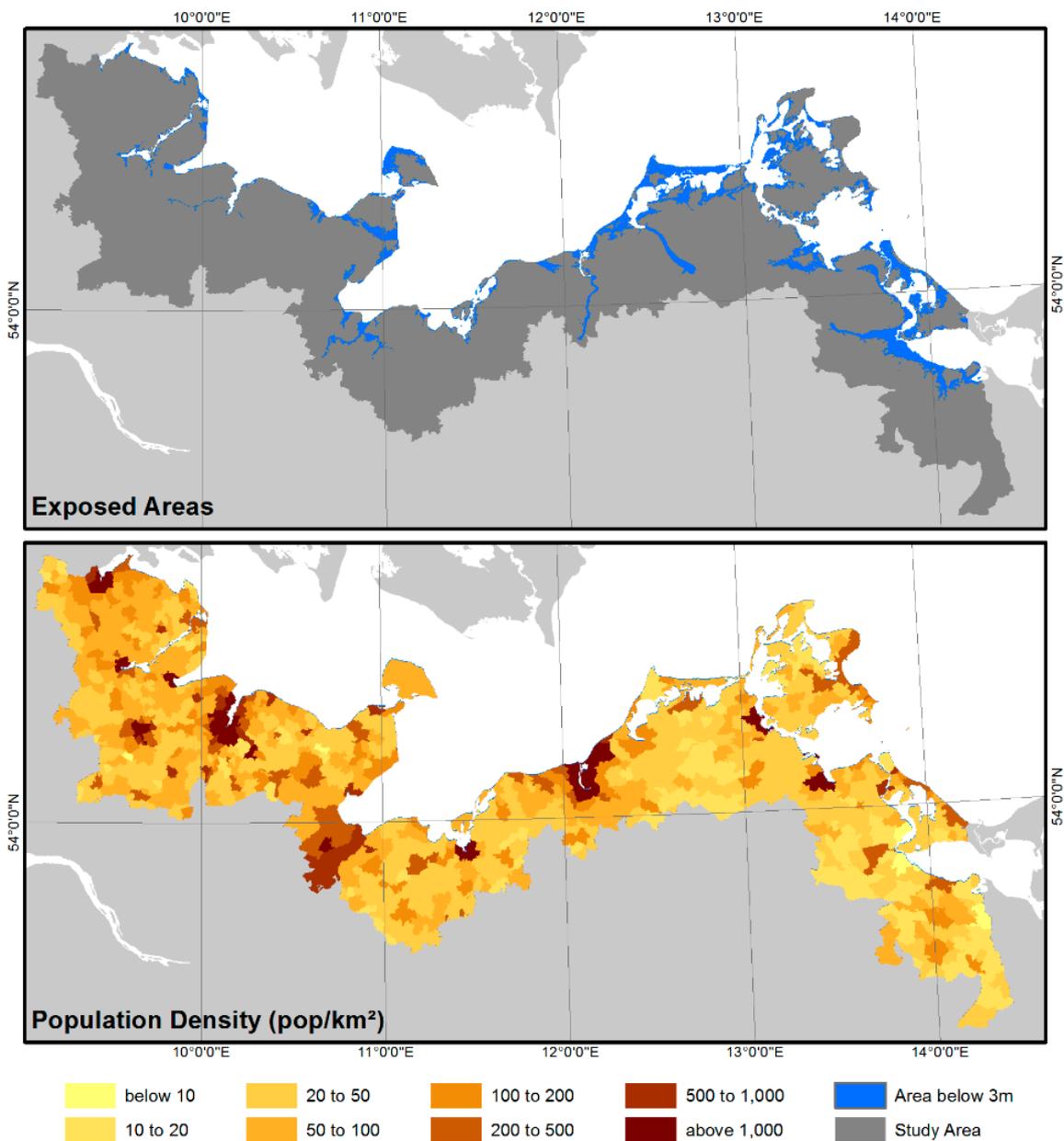


Figure 3. Exposed areas and population density per municipality.

3.2. Resampling of Population

We compared 6 approaches to regionalise population within administrative units (Table 1). Approach I served as the baseline approach presenting population distribution in global datasets. Approaches II to VI employed the GUF. In this context, we used ‘urban’ as areas identified as urban in the GUF, which included cities as well as rural settlements. All approaches were applied on both the district and municipality level.

Table 1. Main concept the 6 approaches tested.

Approach	Description
I	+ no ancillary data used + uniform population density within administrative units
II and III	+ population is only assigned to urban areas + uniform population density within urban areas of an administrative unit
IV	+ population is only assigned to urban areas + population is assigned proportionally to the share of urban extent per cell + uniform population density in cells with the same urban share within one administrative unit + population density across and within settlements of the same administrative unit differ urban share differs
V and VI	+ population is only assigned to urban areas + settlements can extend over more than one administrative unit + population density increases with extent of settlements + population density between settlements of an administrative unit differs + uniform population density within a settlement

In Approach I, we assumed that population is uniformly distributed within each administrative unit, which means that all cells within an administrative unit had the same population count (and density as we used an equal area projection). Between administrative units, population density differed. This approach is similar to the one used in GPW [18] but does not account for water bodies. In Approach II, we identified the urban extent per administrative unit based on GUF2.8. We assumed that population is solely located in these areas and distributed the population uniformly in the urban areas of each administrative unit. All settlements within an administrative unit had the same population density. In Approach III, we used GUF0.4 5% instead of GUF2.8 to identify urban extent per administrative unit and assumed uniform population densities within these areas. Approaches II and III are comparable to dasymetric approaches that distribute population based on land cover or land use (e.g., References [37,38]) but less complex, as they assign population solely to urban areas. In Approach IV, we weighted the population count per administrative unit based on the share of urban extent per 100 m cell calculated from GUF0.4 5%. Within an administrative unit, the population count in a cell with an urban share of 80% was twice as high as in a cell with an urban share of 40%. Between administrative units, the population count of cells with the same urban share differed. In Approach V, we assumed that population density increases with the size of urban clusters. To define urban clusters, we first buffered clusters in GUF2.8 by 150 m (one cell in all directions) to avoid clusters being divided by rivers and parks. Second, we performed an 8-sided connectivity analysis of the buffered GUF2.8 to group cells to clusters. Finally, we reduced the extent of the defined clusters to the original extent of GUF2.8. Next, we calculated the extent and population count for each cluster. We grouped these clusters into small (<10 ha), medium (10 ha to 150 ha), and large (>150 ha) clusters based on the extent. We assumed a linear correlation between the log cluster extent (in ha) and mean population density per cluster. We used the stats package of R Version 3.3.1 [39] to fit linear models for both medium and large clusters. We set the upper threshold to 150 ha to minimize the offset between the linear fit for medium and large clusters and the lower threshold to 10 ha to exclude outliers. For medium clusters, we achieved the best fit with an intercept of -0.771 and a slope of 3.607 . For large clusters, an intercept of -14.16 and a slope of 6.29 led to the best fit. For all small clusters, we applied the modeled population density for an extent of 10 ha. Next, we multiplied the modeled mean population density with the cluster extent to calculate the modeled population count

per cluster. Within a cluster, we assumed uniform population densities. On the administrative level, we adjusted in a final step the modeled population totals to the population totals reported in the census. In Approach VI, we followed the scheme described in Approach V but used GUF0.4_{5%} instead of GUF2.8. Subsequently, we buffered all urban pixels by 150 m. We used the same threshold of 10 ha to differentiate between small and medium clusters but a threshold of 235 ha to differentiate between medium and large clusters. For medium clusters, an intercept of -2.85 and a slope of 2.78 led to the model with the best fit. For large clusters, we achieved the best fit with an intercept of -20.8 and a slope of 6.0 . Approaches V and VI are described in more detail in the Appendix A.

4. Results and Discussion

To assess if the GUF can be used as an estimator for population, we compared the urban extents classified in GUF2.8, in GUF0.4 without any threshold, and in GUF0.4_{5%}. The latter employs an urban coverage threshold of 5% to the areas that actually are populated in the census population raster. Results are presented in Table 2. We found that 83% of the population live in areas that are defined as urban in GUF2.8. 44% of all populated areas are not classified as urban (omission error) and 31% of areas that are classified as urban are not populated at all (commission error). GUF0.4 appears to be a better estimator than GUF 2.8, as 95.3% of the population live in areas classified as urban in GUF0.4. The commission error of 46% is higher as in GUF2.8, but the omission error of 18% is considerably lower. Using a 5% urban coverage threshold reduces the population living in areas classified as urban to 94% and increases the omission error to 22%, but reduces the commission error to 40%. Since the sum of the omission and commission error is lower in GUF0.4_{5%}, we did not use the original GUF0.4 for our analyses.

Table 2. Confusion matrix for population and urban settlements.

Urban Extent	Area (ha) with Population not Classified as Urban	Area (ha) without Population Classified as Urban	Area (ha) with Population Classified as Urban	Omission Error ¹	Commission Error ²	Population Captured ³
GUF2.8	49,960	28,357	63,708	44.0%	30.8%	83.1%
GUF0.4	20,482	79,221	93,186	18.0%	45.9%	95.3%
GUF0.4 _{5%}	25,074	59,183	88,594	22.1%	40.0%	94.1%

¹ Omission error is defined as the share of cells with population >0 that are not classified as urban in the GUF on the total amount of cells with a population greater than zero. ² Commission error is defined as the share of cells that are classified as urban in GUF but with a population of zero on the total amount of cells that are classified as urban in GUF. ³ Population captured is the share of the sum of population in cells that are classified as urban in the GUF on the total sum of population in the study area.

As cells with a population of one have a value of zero in the census population raster, the ‘true’ error of commission is lower than the one reported in Table 2. The percent of population that overlays with urban cells in GUF is also affected by the fact that the values of cells with low population (<3) have been altered in the original census raster [33]. Replacing population counts of 1 by 0 (2 by 3) reduces (increases) the proportion of population in urban areas. As 97.1% of the population is located in cells with a population larger three, we expect this effect to be negligible.

4.1. Performance on Municipality Level

We tested the performance of the six approaches on the district level by comparing the predicted population on municipality level to the census population on the municipality level [24]. The performance metrics of the approaches presented in Table 3 are only representative for the rural districts, as urban districts consist of only one municipality each. In the case that urban districts would consist of more than one municipality, we would expect the differences in performance metrics between the tested approaches to be smaller, as the proportion of built-up area on the total area in urban districts is larger than in rural districts (see Figures 1 and 2). For Approach I, this would lead to an improvement of the performance metrics, as the share of nonurban areas and thus the potential of wrongly allocated population would

reduce. The performance metrics of the other approaches would depend on the actual characteristics of urban areas. We expect all approaches to perform best, if population within cities is homogeneously distributed. In case of heterogeneously distributed population, e.g., due to multistorey housing, which the GUF and consequently our approaches do not consider, we expect the performance of the proposed approaches to decline, as population densities are underestimated. However, if population data are available for city districts, the approaches can resolve differences in population density within cities.

Table 3. Performance metrics for the six approaches ¹.

Approach	GUF	Level of Homogenisation	Q ₂₅ ²	Q ₅₀ ²	Q ₇₅ ²	MAE ³	RTAE ⁴	RMSE ⁵	%RMSE ⁶
I	-	admin level	0	431	867	1278	0.467	2572	94%
II	GUF2.8	urban area	-73	59	257	402	0.147	892	33%
III	GUF0.4 ^{5%}	urban area	0	173	377	570	0.208	1270	46%
IV	GUF0.4 ^{5%}	-	-81	46	218	376	0.137	835	31%
V	GUF2.8	settlement	-280	-121	22	433	0.158	940	34%
VI	GUF0.4 ^{5%}	settlement	-223	-57	89	395	0.144	796	29%

¹ Calculations are based on the difference (error) in total population per municipality between predicted population by the six approaches adjusted to district level minus the census population. ² Q₂₅, Q₅₀ and Q₇₅ are the 25th, 50th, and 75th quantile of the error. ³ MAE is the Mean Absolute Error. ⁴ RTAE is the Relative Total Absolute Error, which is defined as the sum of the deviations between modeled and true population on the municipality level divided by the total population [40]. ⁵ RMSE is the Root Mean Square Error. ⁶ %RMSE is defined as the RMSE divided by the mean population of administrative units [24].

Approach I shows overall the highest differences to the census data. The Mean Average Error (MAE) and the Root Mean Square Error (RMSE) are approximately double of the respective ones in Approach III, which has the second-highest differences to the census data. Both Approaches I and III overestimate population in three out of four municipalities. Approaches II, IV, and VI show similar MAEs, with Approach IV leading to the smallest MAE. Approach VI has the best performance based on RMSE.

Comparing our results, in terms of absolute values, to other studies is not straightforward, as characteristics of settlements in different study areas can considerably differ. In addition, many studies use the RMSE or the MAE to compare model performance within a study area. If different study areas are compared, the information value that these indicators provide is limited, as study areas with high population counts will lead to higher RMSE than study areas with small population counts. To overcome this issue and to provide a first-order comparison, we used the relative indices %RMSE and Relative Total Absolute Error (RTAE); %RMSE is defined as the RMSE divided by the mean population count of the administrative units [24] and RTAE is defined as the sum of the deviations between modeled and true population over all administrative units divided by the total true population [40]. Stevens and colleagues [24] calculated %RMSE for GPW for Cambodia (82%), Vietnam (100%), and Kenya (146%). Approach I, which is comparable to the approach used in GPW, shows an %RMSE of 94%, which agrees with the findings by Stevens and colleagues [24]. The different approaches proposed in this study reduce the %RMSE by a factor of 2 to 3, which is comparable to the factor 1.6 to 2 reported by Stevens and colleagues [24], who use about 40 variables to model population density on the national level. Briggs and colleagues [41] tested satellite data on light emissions and Corine Land Cover data to spatially distribute population for fourteen European countries. They validated their results against four sets of population data on the European level, and for Great Britain separately. For Great Britain, the RMSE ranged from 238 to 346 with reported mean populations of 228 and 370. Based on these reported numbers, we calculated the %RMSE to be 105% and 94%, respectively. The performance is in the same order of magnitude as Approach I in our study, but the error is 2 to 3 times higher than in Approaches II to VI that actually employed the GUF. On the European level, the RMSE in the study of Briggs and colleagues [41] ranged from 240 to 412 with reported mean populations of 228 and 370. This corresponds to an %RMSE of 105% and 114%, respectively. However, Briggs and colleagues validated their results on cell level (1 km resolution), whereas we validated the tested approaches on municipality level. Batista e Silva and

colleagues [40] used a refined version of Corine Land Cover to spatially distribute population data on community level for Europe to a spatial resolution of 100 m. They reported RTAE on national level between 0.106 and 0.892. For Germany, their approaches led to RTAE between 0.247 and 0.281 [40]. The approaches tested in the present study led to an RTAE of 0.467 for Approach I, and a range from 0.137 to 0.208 for the approaches employing GUF data (see Table 3). Based on these numbers, the performance metrics suggest that GUF can lead to slightly better results than land-cover-based approaches. However, our study analyzed a subregion of Germany, whereas the model performance presented by Batista e Silva and colleagues [40] is evaluated nationwide.

The performance of the proposed approaches on the municipality level varies over the study area (Figure 4). Approach I overestimates population in municipalities with mean population densities below the average of the respective district. The overestimation is higher the more concentrated the population is in urban centres. Approaches II and III show in general lower prediction errors than Approach I. However, Approach III considerably overestimates population in some municipalities, which also leads to higher RMSE and %RMSE than in Approach II (Table 3). Approach IV seems to perform well all over the study area and shows the lowest overall MAE. Approach VI and particularly Approach V underestimate population in most municipalities. Furthermore, Approach VI shows overall the lowest RMSE and %RMSE but considerably overestimates population in some municipalities.

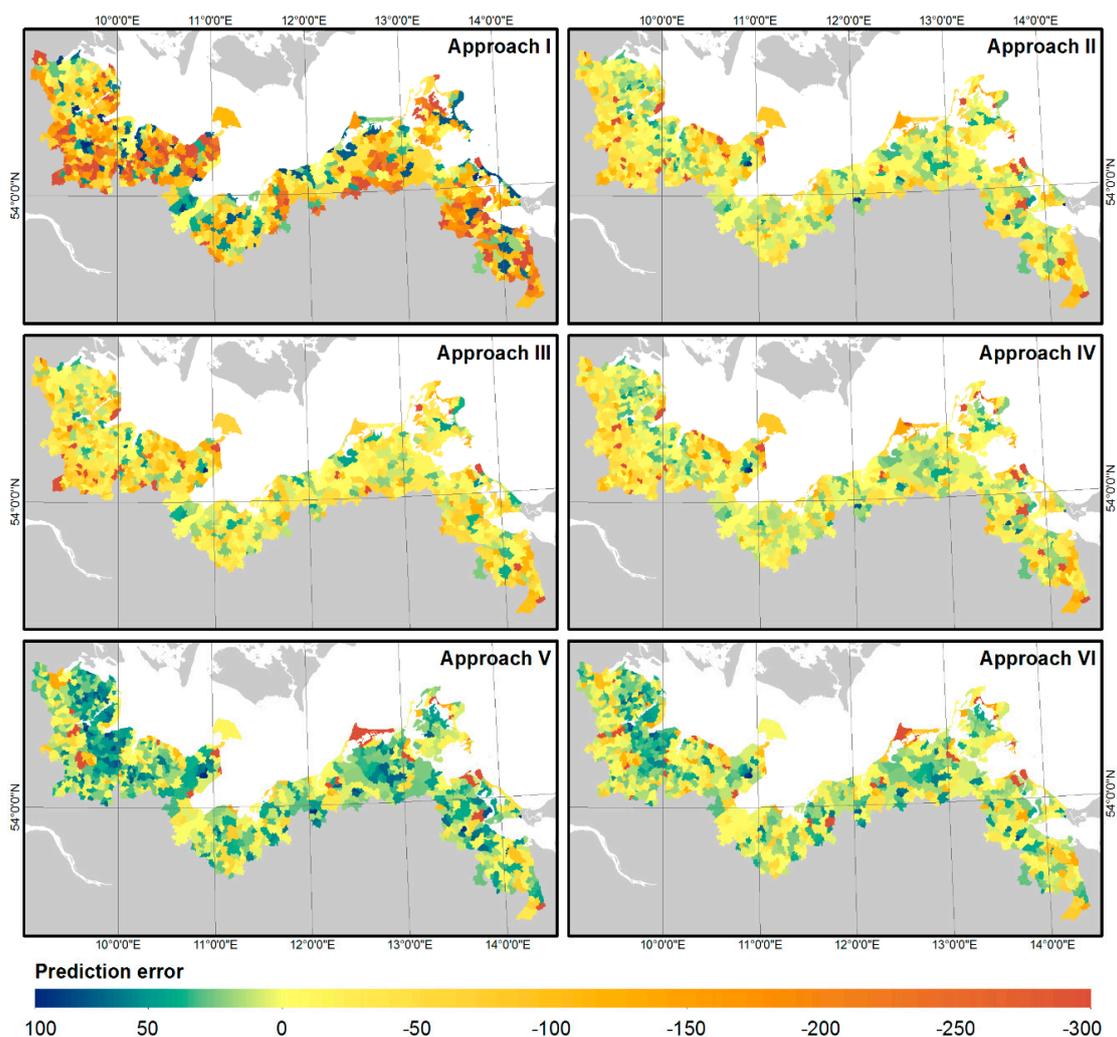


Figure 4. Prediction errors (observed minus predicted) as percentage of the observed population per municipality. Prediction errors bigger than -300 were set to -300 for visualisation.

Approach VI leads in 27.6% of all municipalities to the smallest absolute error (Figure 5), followed by Approach IV (20.5%), Approach V (15.8%), Approach II (14.6%), Approach III (14.2%), and Approach I (5.7%). For 1.6% of the municipalities, at least two approaches led to the smallest error. Approach I leads for 69.3% of all municipalities to the highest absolute error, followed by Approach V (12.6%), Approach VI (7.2%), Approach III (4.9%), Approach II (3.0%), and Approach IV (2.0%). For another 1.0% at least two approaches lead to the highest error. This is in agreement with the results shown in Table 3, which indicate the smallest errors for Approaches VI and IV.

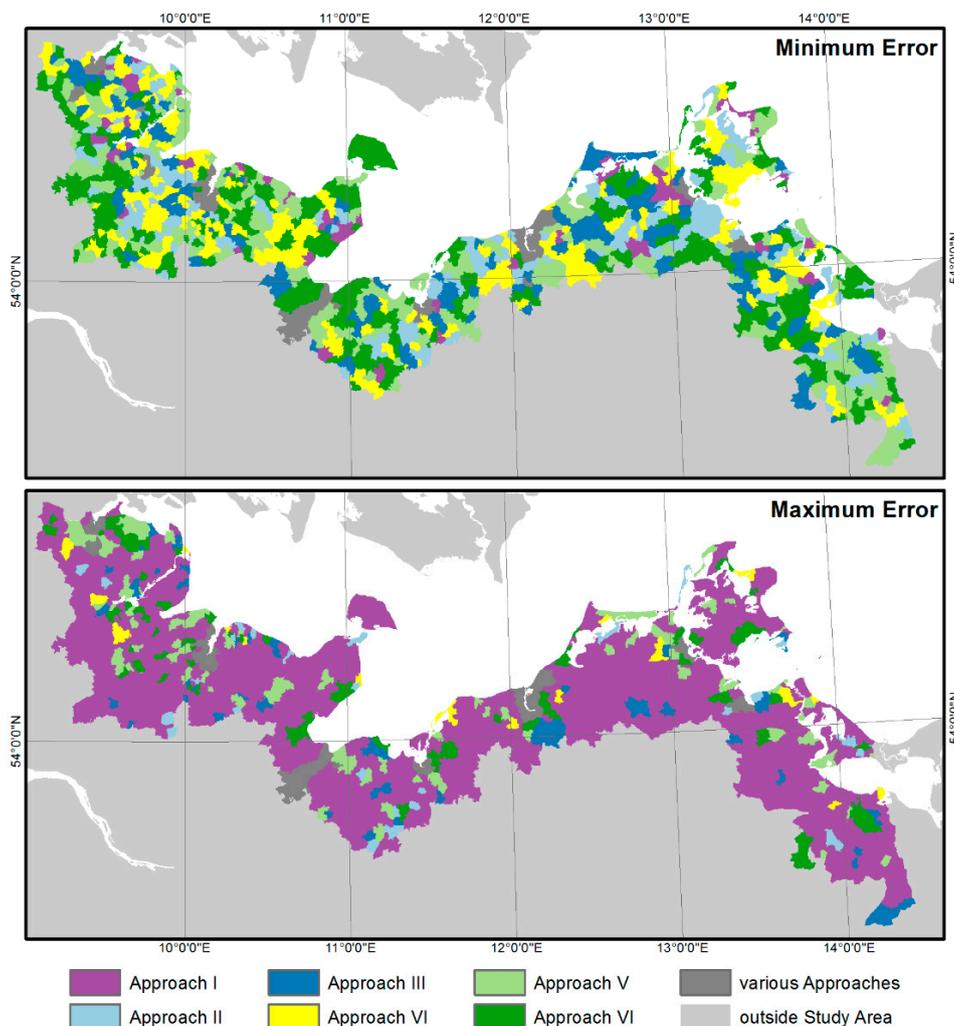


Figure 5. Approaches with the lowest and highest absolute error in predicting population on municipality level.

4.2. Exposed Population

We evaluated the six Approaches by assessing the exposure of population to coastal flooding and comparing it to the ‘true’ exposure. We assessed the ‘true’ exposure by analyzing the population living below 3 m based on the 100 m census grid [32]. We found that all tested approaches led to a considerable overestimation of exposure in our study area. The results are shown in Table 4. Approach V, in which the population density increases with city size, shows the lowest error (overestimation of 72%) if population is adjusted on municipality level. Approach I, which distributes population uniformly on municipality level, shows the highest error (overestimation of 143%).

Table 4. Exposed population for the six approaches adjusted on the district and municipality level.

Approach	GUF	Level of Homogenisation	Level of Adjustment	Pop Exposed	Error	Error%
I	-	admin level	district	218,478	119,043	120
I	-	admin level	municipality	241,293	141,858	143
II	GUF2.8	urban area	district	185,517	86,082	87
II	GUF2.8	urban area	municipality	174,280	74,856	75
III	GUF0.4 5%	urban area	district	188,853	89,418	90
III	GUF0.4 5%	urban area	municipality	183,793	84,358	85
IV	GUF0.4 5%	-	district	187,490	88,055	89
IV	GUF0.4 5%	-	municipality	174,465	75,030	75
V	GUF2.8	settlement	district	184,052	84,617	85
V	GUF2.8	settlement	municipality	170,623	71,189	72
VI	GUF0.4 5%	settlement	district	191,590	92,155	93
VI	GUF0.4 5%	settlement	municipality	182,035	82,600	83
		'True' Exposure		99,435		

For all approaches that employ the GUF, we found that population adjustments on finer levels lead to improved estimates in exposure analysis. As population is not distributed uniformly within an administrative unit, errors are reduced by adjusting data on finer administrative units. Nevertheless, in Approach I, which is comparable to widely used approaches on global and regional scales, the adjustment of the data on finer scale led to considerably higher errors in exposed population. This suggests that, in the German Baltic Sea region, people tend to live near the coast but not in exposed areas. For example, the population density in coastal rural municipalities is 34% higher than in non-coastal rural municipalities. If the seven municipalities of urban districts (which are all coastal) are also considered, population density in coastal municipalities is 175% higher than in non-coastal municipalities.

We also found that previous studies were likely to have overestimated exposed population, although different reference dates for population and the modeled flood heights make a direct comparison difficult. For example, Sterr [30] used population numbers on the municipality and district level to estimate the 1995 population living on an elevation of 5 m or lower above mean sea level for Mecklenburg-Vorpommern at 319,400. This value is more than three times higher than the 2011 exposure of population living on elevations of 3 m or lower above mean sea level reported for the entire study area. Considering a decline of population in Mecklenburg-Vorpommern of 12% between 1995 and 2011 [42], the exposed population is in the same magnitude as in Approach I or even higher.

Our results demonstrate that the use of the GUF reduces the overestimation of exposed population by 22% to 27% on district level and by 40% to 50% on municipality level (Table 4). This shows that by not assigning any population to areas with no urban settlements errors in impact assessment can be reduced considerably. We also found that approaches with GUF2.8 led to better results than approaches with GUF0.4 5% (see Approach II vs. Approach III, and Approach V vs. Approach VI). This may seem surprising, as GUF2.8 covers 83% and GUF0.4 5% 94% of the population in the study area. However, according to our findings the commission error in GUF0.4 5% is higher than in GUF2.8 (see Table 2), which means that more uninhabited areas are classified as urban. One reason for this is that urban uninhabited structures, such as industrial areas, secondary housing, or small buildings, are resolved at finer spatial resolutions. Approach IV shows that GUF0.4 5% can lead to improved results, if it is used to differentiate population density within settlements.

We must note that our study is based on census data. This means that the population counts represent where the population is registered but not necessarily where the population resides. This is of particular relevance for exposure analysis. Deville and colleagues [20] show for France and Portugal, that during the summer holiday period (July and August), coastal areas face a population increase of > 60% compared to the population officially registered in the coastal areas. We assume that their findings also hold true for the German Baltic coast. As surges in the study area are observed during winter, when few tourists are in the study area, we did not explicitly account for tourism. However, as GUF does not differentiate between touristic accommodation (e.g., secondary houses and hotels) and houses

permanently inhabited, we accounted implicitly for tourism. This also led to an overestimation of exposed population in all Approaches II to VI, if touristic infrastructure is located in exposed areas and residential population are assigned erroneously to these areas. The same effect can be seen in industrial areas. These are also resolved in GUF but have very low population density or are uninhabited. In the study area, this led to an overestimation of exposure, as harbors and shipyards are typically located on low elevated zones close to the coastline. A further limitation is that large agglomerations of people, e.g., multistorey residential buildings, cannot be represented by all tested approaches. This can lead to considerable underestimation of population in a small number of cells. This limitation does not only apply to this study but, in general, to dasymetric-mapping approaches [41].

We expect that our findings are transferable to other study regions. Globally, 39% of the population lives within a distance of 100 km from the coast [43]. This population is also not distributed uniformly within administrative units but gathered in urban clusters [43]. As the GUF seems to provide realistic settlement patterns in study areas all over the globe [27], we expect that the proposed approaches II to IV can be applied on regional and global scales to regionalize population within administrative units, which could considerably improve the data basis for exposure analysis. Approaches V and VI perform well (Approach V lowest overall error for exposure; Approach VI lowest overall %RMSE) but rely on highly resolved census data to adjust the thresholds used in the modeling process. These are not available all over the globe [44,45]. Approach II performs better than Approach III as GUF0.4 5% has a higher error of commission than GUF2.8. However, GUF2.8 does not capture 17% of the total population within the study area. Weighting the population depending on the urban coverage in each cell (Approach IV) leads to the smallest errors on municipality level and reduces the error in exposure analysis considerable.

5. Conclusions

Our study shows that using uniform population densities on a municipality level (finer scale) can lead to higher errors in exposure analysis compared to using uniform population densities on a district level (coarser scale). As new population data tend to be available on ever finer scales [18], the assumption of a uniform spatial distribution of population when assessing exposure to coastal flooding may lead to substantial errors in assessing exposure to coastal flooding. By using simple methods that solely employ the GUF as ancillary data to regionalise population within administrative units the error in exposed population can be reduced by 40% to 50%. However, exposure analysis shows that the modeled population distributions overestimate the number of people living in the floodplain compared to gridded census data. We anticipate that accounting for the height of buildings in order to determine the number of floors can lead to improved estimates. Before any of the analyzed approaches can be applied on global scales, the observations of this study would need to be further evaluated in study areas with different settlement characteristics.

Author Contributions: Conceptualization, J.-L.M. and A.T.V.; methodology, J.-L.M.; formal analysis, J.-L.M.; investigation, J.-L.M.; writing—original draft preparation, J.-L.M.; writing—review and editing, A.T.V.

Funding: This research was funded by the German Research Foundation (DFG) Priority Program (SPP) 1889 Regional Sea Level Change and Society (SeaLevel).

Acknowledgments: The authors would like to thank the Mecklenburg-Vorpommern office of internal administration and the Schleswig-Holstein land surveying and geoinformation office for providing elevation data. We would also like to express our thanks to the editor and the reviewers for their valuable comments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix

Model Description

We used nine (four urban and five rural) districts to calibrate the model used in Approaches V and VI. For validation, we used eight (three urban and five rural) districts. The districts were selected

randomly, but followed the conditions that two urban and two or three rural districts of each state had to be used for calibration. These conditions ensured that cities could be found in both calibration and validation data. Furthermore, the conditions allowed accounting for possible differences in settlement patterns that developed under different political systems in the study area between 1945 and 1989 [46] and might persist.

Approach V

In Approach V, we used 3043 settlements (urban clusters) that were located in the districts selected for calibration. Of these, we classified 2352 settlement as small clusters (extent < 10 ha), 653 as medium clusters (extent \geq 10 ha and <150 ha), and 38 as large clusters (extent \geq 150 ha). For medium and large clusters, we used a linear model to represent the correlation between the logarithm of settlement extent and the mean population density per settlement calculated by the stats package in R version 3.3.1 [39]. For small clusters, we used the modeled density for a settlement extent of 10 ha, independent from the actual settlement extent. For medium clusters, we achieved the best fit with an intercept of -0.771 and a slope of 3.607 . For large clusters, an intercept of -14.16 and a slope of 6.29 lead to the best fit. For validation, we used 1911 settlements located in the districts selected for validation of which we classified 1336 as small clusters, 533 as medium clusters and 42 as large clusters.

We evaluated the performance of the model by calculating the RMSE and the MAE for mean population density within a settlement and the sum of population in a settlement (Table A1).

Table A1. Model performance in Approach V (using GUF2.8).

Parameter	Calibration Mean Density	Validation Mean Density	Calibration Population Sum	Validation Population Sum
RMSE	5.6	5.7	1007	854
MAE	4.7	4.7	84	96

Figure A1 illustrates the model used in Approach V and the settlement characteristics (extent and mean population density) used for calibration and validation.

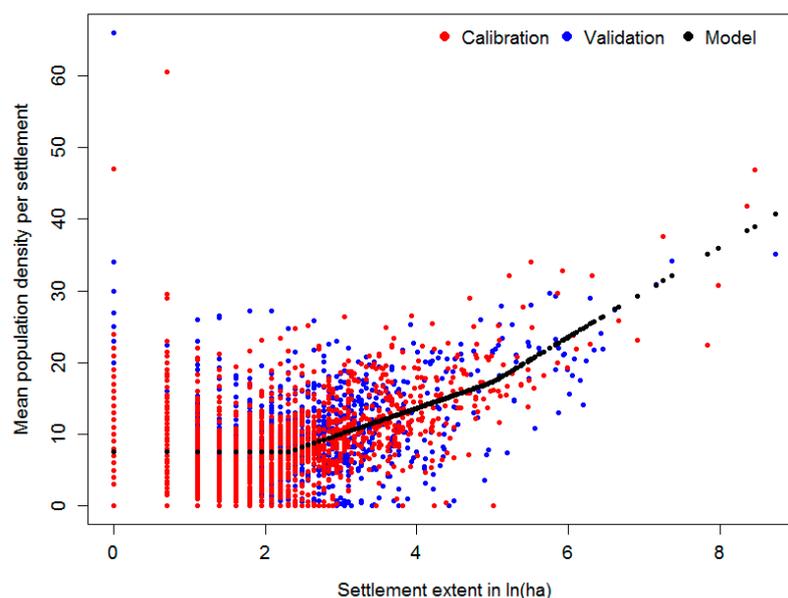


Figure A1. Observed and modelled mean population density and settlement extent for calibration and validation in Approach V.

GUF2.8 captures 83.1% of the population in the study area. As the model solely used GUF2.8 as ancillary data, this led to an underestimation of the total population in the study area. Thus, we adjusted the predicted population counts to the census population on the district and municipality level [28]. This led to an increase of RMSE and MAE for mean population density within a settlement (Table A2), because people actually not living in areas defined as settlements by GUF2.8 were located to areas defined as settlements. Despite this, the RMSE of the population sum per settlement adjusted to municipality level was reduced considerably.

Table A2. Model performance of adjusted model in Approach V.

Parameter	Calibration Mean Density	Validation Mean Density	Calibration Population Sum	Validation Population Sum
RMSE (d ¹)	6.6	6.4	1025	682
MAE (d ¹)	5.7	5.3	95	101
RMSE (m ²)	11.1	10.7	537	419
MAE (m ²)	9.1	7.9	85	90

¹ adjusted to total population on district level. ² adjusted to total population on municipality level.

Figure A2 shows the residuals for the unadjusted model, the model adjusted to match total district population and the model adjusted to match total municipality population. The adjusted models overestimate (positive residuals) the mean population density for large clusters. Furthermore, the model adjusted to the municipality level shows high residuals for very small clusters, which indicates a small number of urban cells in the corresponding municipality.

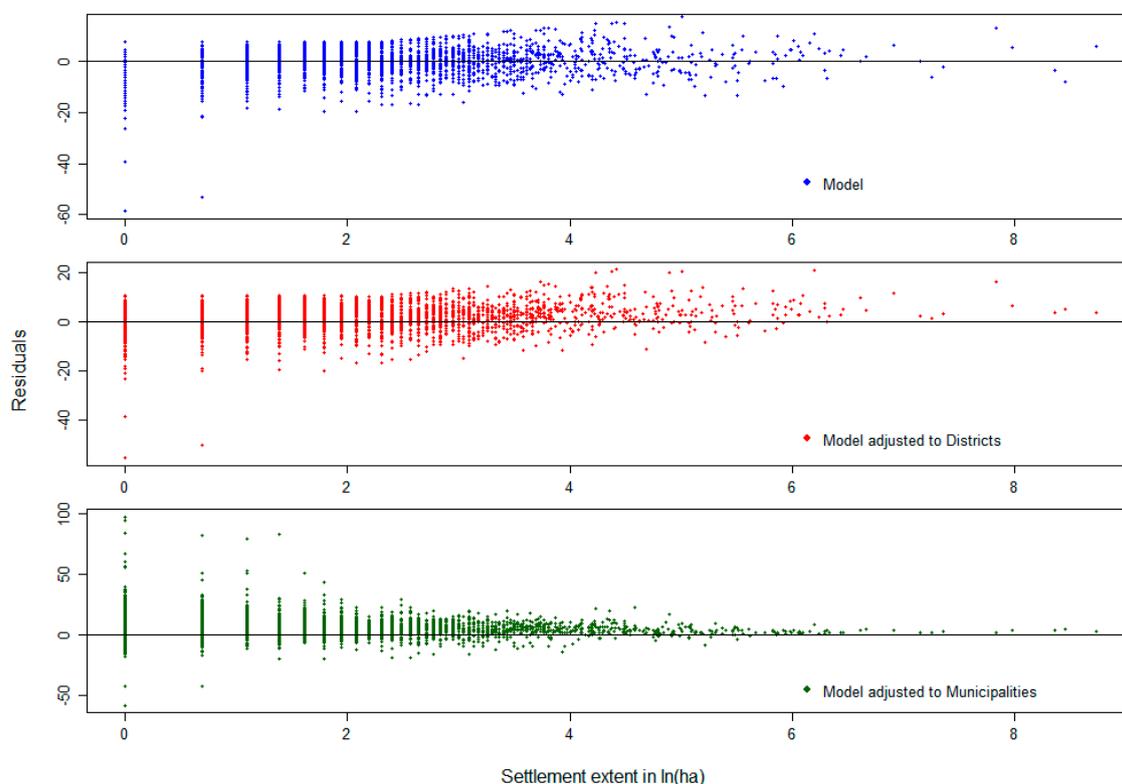


Figure A2. Residuals (modelled mean population density minus observed mean population density) and settlement extent in Approach V.

Approach VI

In Approach VI, we used the same districts for calibration and validation as we did for Approach V. Due to different urban extents the number of settlements (urban clusters) differed between the approaches. We grouped 3232 settlements in the calibration area to 2057 small clusters (below 10 ha settlement extent), 1136 medium clusters (extent ≥ 10 ha and <235 ha), and 39 large clusters (extent ≥ 235 ha). For medium clusters, a model with an intercept of -2.85 and a slope of 2.78 showed the best fit. For large clusters, an intercept of -20.8 and a slope of 6.0 led to the model with the best fit. We used the calculated mean population density for clusters with an extent of 10 ha for small clusters independently from the actual cluster extent. For validation, we grouped 2047 settlements to 1177 small clusters, 829 medium clusters, and 41 large clusters. We calculated RMSE and MAE to test the ability of the model to predict mean population density per settlement based on the settlement extent (Table A3).

Table A3. Model performance for Approach VI (using GUF0.4_{5%}).

Parameter	Calibration Mean Density	Validation Mean Density	Calibration Population Sum	Validation Population Sum
RMSE	3.2	3.6	918	1077
MAE	2.4	2.6	87	107

Figure A3 illustrates the model used in Approach VI and the settlement characteristics (extent and mean population density) used for calibration and validation.

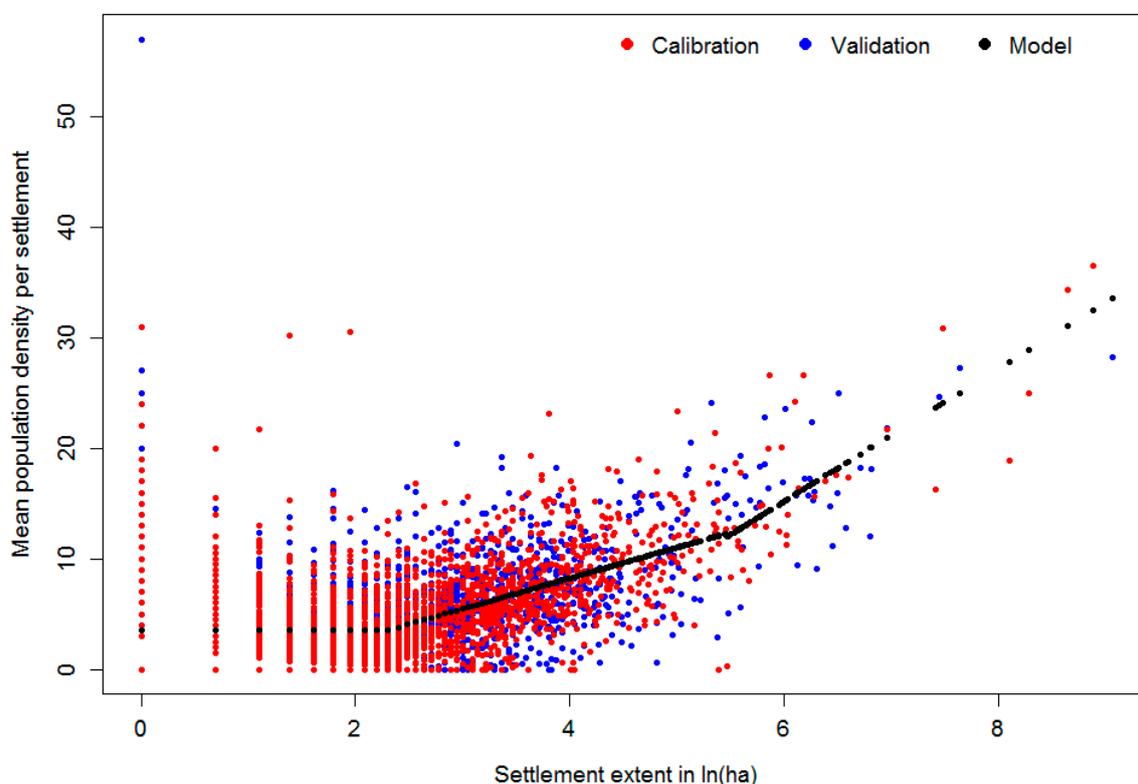


Figure A3. Observed and modelled mean population density and settlement extent for calibration and validation in Approach VI.

GUF0.4_{5%} captured 94.1% of the population in the study area, which is clearly more than GUF2.8 (see Table 2). This required smaller adjustments on district and municipality level compared to

Approach V, which led to smaller errors in the model performance of the adjusted model (Table A4). The error indicators for population sum in calibration and validation areas were reduced considerably compared to Table A3.

Table A4. Model performance of adjusted model in Approach VI.

Parameter	Calibration Mean Density	Validation Mean Density	Calibration Population Sum	Validation Population Sum
RMSE (d ¹)	3.2	3.6	611	447
MAE (d ¹)	2.5	2.7	71	86
RMSE (m ²)	3.5	4.2	171	198
MAE (m ²)	2.7	2.9	43	55

¹ adjusted to total population on district level. ² adjusted to total population on municipality level.

Comparable to Approach V (Figure A2), the adjusted models overestimated (positive residuals) the actual mean population density in large clusters (Figure A4). However, in particular the adjustment to population totals on the municipality level led to a reduction of residuals for large clusters and lowered the RMSE by factor 5 (171 (calibration) and 198 (validation) compared to 918 (calibration) and 1077 (validation) in the unadjusted model).

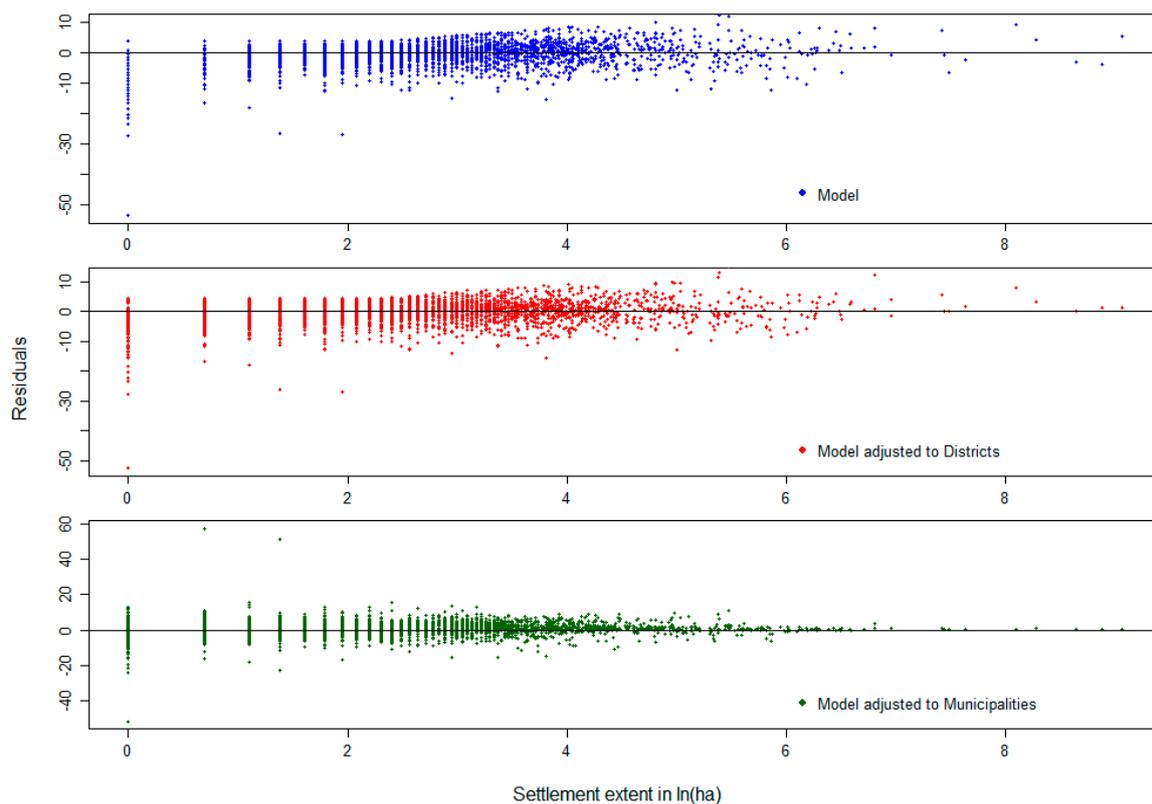


Figure A4. Residuals (modeled mean population density minus observed mean population density) and settlement extent in Approach VI.

References

1. Kron, W. Coasts: The high-risk areas of the world. *Nat. Hazards* **2013**, *66*, 1363–1382. [[CrossRef](#)]
2. Nicholls, R.J.; Cazenave, A. Sea-level rise and its impact on coastal zones. *Science* **2010**, *328*, 1517–1520. [[CrossRef](#)] [[PubMed](#)]

3. Wong, P.P.; Losada, I.J.; Gattuso, J.-P.; Hinkel, J.; Khattabi, A.; McInnes, K.L.; Saito, Y.; Sallenger, A. Coastal systems and low-lying areas. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability: Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., et al., Eds.; Cambridge University Press: Cambridge, UK, 2014; pp. 361–409.
4. Hinkel, J.; Lincke, D.; Vafeidis, A.T.; Perrette, M.; Nicholls, R.J.; Tol, R.S.J.; Marzeion, B.; Fettweis, X.; Ionescu, C.; Levermann, A. Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3292–3297. [[CrossRef](#)] [[PubMed](#)]
5. Neumann, B.; Vafeidis, A.T.; Zimmermann, J.; Nicholls, R.J. Future coastal population growth and exposure to sea-level rise and coastal flooding—A global assessment. *PLoS ONE* **2015**, *10*, e0118571. [[CrossRef](#)] [[PubMed](#)]
6. Merkens, J.-L.; Reimann, L.; Hinkel, J.; Vafeidis, A.T. Gridded population projections for the coastal zone under the Shared Socioeconomic Pathways. *Glob. Planet. Chang.* **2016**, *145*, 57–66. [[CrossRef](#)]
7. Jones, B.; O'Neill, B.C. Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. *Environ. Res. Lett.* **2016**, *11*, 084003. [[CrossRef](#)]
8. Brown, S.; Nicholls, R.J.; Woodroffe, C.D.; Hanson, S.; Hinkel, J.; Kebede, A.S.; Neumann, B.; Vafeidis, A.T. Sea-Level Rise Impacts and Responses: A Global Perspective. In *Coastal Hazards*; Finkl, C.W., Ed.; Springer: Dordrecht, The Netherlands, 2013; pp. 117–149.
9. Neumann, B.; Ott, K.; Kenchington, R. Strong sustainability in coastal areas: A conceptual interpretation of SDG 14. *Sustain. Sci.* **2017**, *12*, 1019–1035. [[CrossRef](#)] [[PubMed](#)]
10. Voudoukas, M.I.; Voukouvalas, E.; Mentaschi, L.; Dottori, F.; Giardino, A.; Bouziotas, D.; Bianchi, A.; Salamon, P.; Feyen, L. Developments in large-scale coastal flood hazard mapping. *Nat. Hazards Earth Syst. Sci.* **2016**, *16*, 1841–1853. [[CrossRef](#)]
11. Crowell, M.; Coulton, K.; Johnson, C.; Westcott, J.; Bellomo, D.; Edelman, S.; Hirsch, E. An Estimate of the U.S. Population Living in 100-Year Coastal Flood Hazard Areas. *J. Coast. Res.* **2010**, *262*, 201–211. [[CrossRef](#)]
12. Hallegatte, S.; Green, C.; Nicholls, R.J.; Corfee-Morlot, J. Future flood losses in major coastal cities. *Nat. Clim. Chang.* **2013**, *3*, 802–806. [[CrossRef](#)]
13. Center for International Earth Science Information Network—Columbia University (CIESIN). *Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 10*; NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY, USA, 2017.
14. Center for International Earth Science Information Network—Columbia University (CIESIN); International Food Policy Research Institute (IFPRI); The World Bank; Centro Internacional de Agricultura Tropical (CIAT). *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Count Grid*; NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY, USA, 2015.
15. Paprotny, D.; Morales-Nápoles, O.; Jonkman, S.N. HANZE: A pan-European database of exposure to natural hazards and damaging historical floods since 1870. *Earth Syst. Sci. Data* **2018**, *10*, 565–581. [[CrossRef](#)]
16. Reimann, L.; Merkens, J.-L.; Vafeidis, A.T. Regionalized Shared Socioeconomic Pathways: Narratives and spatial population projections for the Mediterranean coastal zone. *Reg. Environ. Chang.* **2018**, *18*, 235–245. [[CrossRef](#)]
17. McGranahan, G.; Balk, D.; Anderson, B. The rising tide: Assessing the risks of climate change and human settlements in low elevation coastal zones. *Environ. Urban.* **2007**, *19*, 17–37. [[CrossRef](#)]
18. Doxsey-Whitfield, E.; MacManus, K.; Adamo, S.B.; Pistolesi, L.; Squires, J.; Borkovska, O.; Baptista, S.R. Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Pap. Appl. Geogr.* **2015**, *1*, 226–234. [[CrossRef](#)]
19. Balk, D.L.; Deichmann, U.; Yetman, G.; Pozzi, F.; Hay, S.I.; Nelson, A. Determining Global Population Distribution: Methods, Applications and Data. In *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications*; Meltzer, M.I., Ed.; Elsevier: New York City, NY, USA, 2006; pp. 119–156.
20. Deville, P.; Linard, C.; Martin, S.; Gilbert, M.; Stevens, F.R.; Gaughan, A.E.; Blondel, V.D.; Tatem, A.J. Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15888–15893. [[CrossRef](#)] [[PubMed](#)]

21. Sorichetta, A.; Hornby, G.M.; Stevens, F.R.; Gaughan, A.E.; Linard, C.; Tatem, A.J. High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Sci. Data* **2015**, *2*, 150045. [[CrossRef](#)] [[PubMed](#)]
22. Gaughan, A.E.; Stevens, F.R.; Linard, C.; Jia, P.; Tatem, A.J. High resolution population distribution maps for Southeast Asia in 2010 and 2015. *PLoS ONE* **2013**, *8*, e55882. [[CrossRef](#)] [[PubMed](#)]
23. Linard, C.; Gilbert, M.; Snow, R.W.; Noor, A.M.; Tatem, A.J. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE* **2012**, *7*, e31743. [[CrossRef](#)] [[PubMed](#)]
24. Stevens, F.R.; Gaughan, A.E.; Linard, C.; Tatem, A.J. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* **2015**, *10*, e0107042. [[CrossRef](#)] [[PubMed](#)]
25. Voudoukas, M.I.; Mentaschi, L.; Voukouvalas, E.; Bianchi, A.; Dottori, F.; Feyen, L. Climatic and socioeconomic controls of future coastal flood risk in Europe. *Nat. Clim. Chang.* **2018**, *3*, 802. [[CrossRef](#)]
26. Esch, T.; Schenk, A.; Ullmann, T.; Thiel, M.; Roth, A.; Dech, S. Characterization of Land Cover Types in TerraSAR-X Images by Combined Analysis of Speckle Statistics and Intensity Information. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1911–1925. [[CrossRef](#)]
27. Esch, T.; Heldens, W.; Hirner, A.; Keil, M.; Marconcini, M.; Roth, A.; Zeidler, J.; Dech, S.; Strano, E. Breaking new ground in mapping human settlements from space—The Global Urban Footprint. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 30–42. [[CrossRef](#)]
28. Federal Statistical Office of Germany, 2011 Census. Census Database of the Census 2011. Available online: <https://ergebnisse.zensus2011.de/?locale=en> (accessed on 4 June 2018).
29. Ministry of Agriculture and the Environment Mecklenburg-Vorpommern (MLUV-MV). *Regelwerk Küstenschutz Mecklenburg-Vorpommern: Übersichtsherft. Grundlagen, Grundsätze, Standortbestimmung und Ausblick*; Ministry of Agriculture and the Environment Mecklenburg-Vorpommern (MLUV-MV): Schwerin, Germany, 2009. (In German)
30. Sterr, H. Assessment of Vulnerability and Adaptation to Sea-Level Rise for the Coastal Zone of Germany. *J. Coast. Res.* **2008**, *242*, 380–393. [[CrossRef](#)]
31. Ministry of Energy, Agriculture, the Environment, Nature and Digitalization Schleswig-Holstein (MELUND-SH). *Generalplan Küstenschutz des Landes Schleswig-Holstein: Fortschreibung 2012*; Ministry of Energy, Agriculture, the Environment, Nature and Digitalization Schleswig-Holstein (MELUND-SH): Kiel, Germany, 2013. (In German)
32. Federal Statistical Office of Germany, 2011 Census. Bevölkerung. Available online: https://www.destatis.de/DE/Methoden/Zensus_/Downloads/csv_Bevoelkerung.zip?__blob=publicationFile (accessed on 4 June 2018).
33. Federal Statistical Office of Germany, 2011 Census. User Information on SAFE. Available online: https://www.zensus2011.de/SharedDocs/Downloads/EN/Publications/information_material/User_information_on_SAFE.pdf?__blob=publicationFile&v=8 (accessed on 13 July 2018).
34. Federal Statistical Office of Germany, 2011 Census. Special evaluation, Results of the Census of 9 May 2011 per grid cell. Available online: https://www.zensus2011.de/SharedDocs/Downloads/DE/Pressemitteilung/DemografischeGrunddaten/ExplanatoryNotes_100m_Population.pdf?__blob=publicationFile&v=3 (accessed on 13 July 2018).
35. Working Committee of the Surveying Authorities of the Laender of the Federal Republic of Germany (AdV). Produktdatenblatt, Digitales Geländemodell Gitterweite 1 m (DGM1). Available online: <http://www.adv-online.de/AdV-Produkte/Standards-und-Produktblaetter/Produktblaetter/binarywriterservlet?imgUid=15653624-758e-6212-df2d-788a438ad1b2&uBasVariant=11111111-1111-1111-1111-111111111111> (accessed on 14 July 2018).
36. Poulter, B.; Halpin, P.N. Raster modelling of coastal flooding from sea-level rise. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 167–182. [[CrossRef](#)]
37. Gallego, F.J. A population density grid of the European Union. *Popul. Environ.* **2010**, *31*, 460–473. [[CrossRef](#)]
38. Linard, C.; Gilbert, M.; Tatem, A.J. Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal* **2011**, *76*, 525–538. [[CrossRef](#)] [[PubMed](#)]
39. R Core Team. *R: A Language and Environment for Statistical Computing*; The R Development Core Team: Vienna, Austria, 2016.

40. Batista e Silva, F.; Gallego, J.; Lavalle, C. A high-resolution population grid map for Europe. *J. Maps* **2013**, *9*, 16–28. [[CrossRef](#)]
41. Briggs, D.J.; Gulliver, J.; Fecht, D.; Vienneau, D.M. Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote Sens. Environ.* **2007**, *108*, 451–466. [[CrossRef](#)]
42. Statistisches Amt Mecklenburg-Vorpommern (StatA MV). *Statistisches Jahrbuch Mecklenburg-Vorpommern 2017*; StatA MV: Schwerin, Germany, 2017. (In German)
43. Kumm, M.; De Moel, H.; Salvucci, G.; Viviroli, D.; Ward, P.J.; Varis, O. Over the hills and further away from coast: Global geospatial patterns of human and environment over the 20th–21st centuries. *Environ. Res. Lett.* **2016**, *11*, 34010. [[CrossRef](#)]
44. Wardrop, N.A.; Jochem, W.C.; Bird, T.J.; Chamberlain, H.R.; Clarke, D.; Kerr, D.; Bengtsson, L.; Juran, S.; Seaman, V.; Tatem, A.J. Spatially disaggregated population estimates in the absence of national population and housing census data. *Proc. Nat. Acad. Sci. USA* **2018**, *115*, 3529–3537. [[CrossRef](#)] [[PubMed](#)]
45. Tatem, A.J.; Campiz, N.; Gething, P.W.; Snow, R.W.; Linard, C. The effects of spatial population dataset choice on estimates of population at risk of disease. *Pop. Health Metrics* **2011**, *9*, 4. [[CrossRef](#)] [[PubMed](#)]
46. Berentsen, W.H. Changing settlement patterns in the German Democratic Republic: 1945–1976. *Geoforum* **1982**, *13*, 327–337. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).