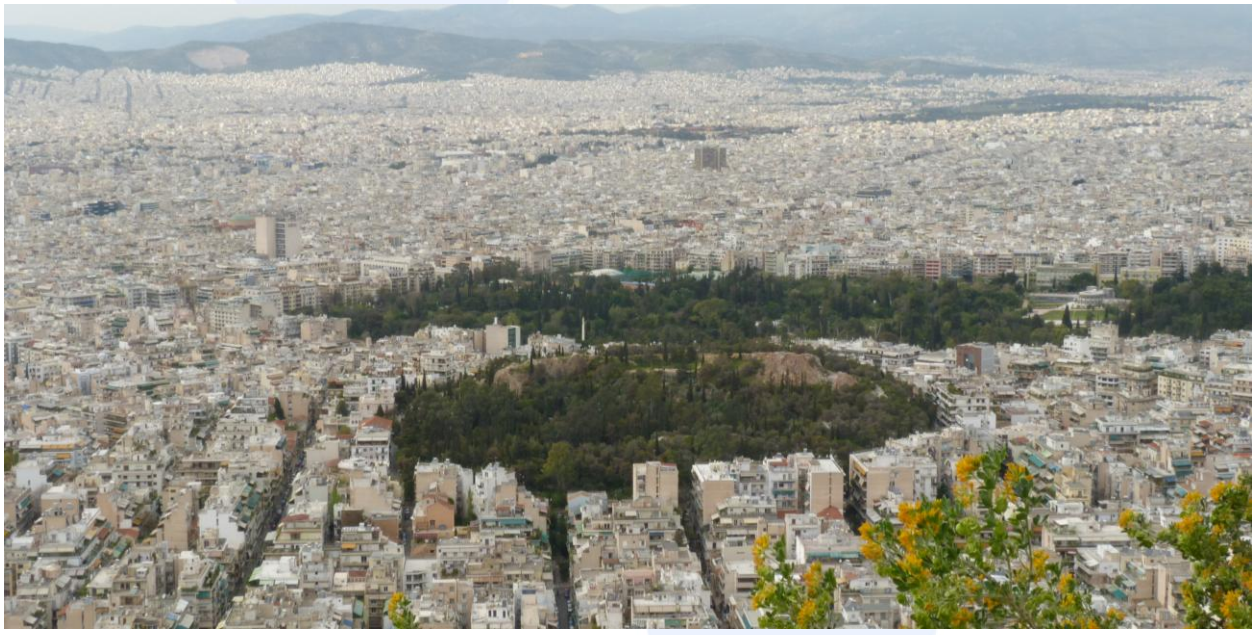
European Commission

Joint Research Centre

Introducing the CRISP model to downscale future population projections

Jacobs-Crisioni, C., Schiavina, M., Krasnodebska, K., Dijkstra, L., Claassens, J., Hilferink, M., van Der Wielen, T., Koomen, E.

2025



This document is a publication by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The contents of this publication do not necessarily reflect the position or opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Name: Lewis Dijkstra
Address: Via E. Fermi 2749, Ispra, Italy
Email: lewis.dijkstra@ec.europa.eu

The Joint Research Centre: EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC143916
EUR 40499

PDF ISBN 978-92-68-32716-6 ISSN 1831-9424 doi:10.2760/7163875 KJ-01-25-529-EN-N

Luxembourg: Publications Office of the European Union, 2025

© European Union, 2025

The reuse policy of the European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of photos or other material that is not owned by the European Union permission must be sought directly from the copyright holders.

How to cite this report: Jacobs-Crisioni, C., Schiavina, M., Krasnodebska, K., Dijkstra, L., Claassens, J. et al., *Introducing the CRISP model to downscale future population projections,* Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/7163875, JRC143916.

Contents

Abstract

The CRISP model has been setup to disaggregate national population projections to 1km^2 grid cells globally from 2020 to 2100. This is done primarily to provide projections of population by degree of urbanisation for the 2025 World Urbanisation Prospects report produced by the UN. Many other applications are feasible with CRISP as well. The model estimates population and built-up area change in a three-step process. First, population and built-up area change are estimated for roughly 1000 functional areas taking into account national population projections. Second, new built-up area is allocated to grid cells considering distance to settlements, roads, water, current share of built-up area and other characteristics. Finally, population is added to newly built-up areas and more suitable locations and reduced in less suitable locations to capture internal migration (and natural population decline).

Acknowledgements

The CRISP model is a further development of the 2UP model developed by the Netherlands Environmental Assessment. We are grateful for their willingness to share their model and data, and would like to thank Bas van Bemmel and Frank van Rijn in particular for their help in setting up and developing CRISP.

Authors

Chris Jacobs-Crisioni, Marcello Schiavina, Katarzyna Krasnodębska, Lewis Dijkstra, Jip Claassens, Maarten Hilferink, Thijmen van der Wielen, Eric Koomen

Executive summary

This report presents the CRISP (Cities and Rural Integrated Spatial Projections) model, a functional method by which national projections of population change are downscaled to the 1 km grid level. This is done by first disaggregating national projections to intermediate units where countries are large. In the initial application, such intermediate units are Functional Areas (FAs). Those FAs were defined for this study specifically with the aim of providing cohesive subnational units of approximately comparable sizes. After downscaling population to the FA level, built-up surface projections at the FA-level are computed using expected changes in built-up surface per capita as an intermediary.

Subsequently, regional built-up surface projections up to year 2100 are downscaled to gridded, 1 km resolution built-up fractions in a statistically informed approach. Locational attractiveness in the built-up model is informed by parameters estimated in logit models. In addition, a combination of rebalancing and an empirically derived function of expected built-up development is used to make sure that the model allocates built-up development in attractive locations, while maintaining persistent built-up surface distributions.

The modelled built-up fractions are subsequently used to guide the allocation of population change. Subsequently population change is modelled. To this end first creating pools of population at the FA level. These pools contain projected net population change and a small percentage of internal migrants that are drawn out of grid cells in the FA. If the population in a FA is expected to decline, a larger share of the population is drawn out of grid cells to ensure that the model can allocate some internal movement. Migrants are expected to move predominantly from grid cells that are considered unattractive.

Population is subsequently allocated from the FA-pool to grid cells using universal rules that reflect stylized assumptions and observed maximum changes, which act as limits to local population change. The population model is driven by a set of assumptions out of necessity. Statistical models of global local population change are still evasive, as the available requisite population grids are a product of dasymetric mapping methods and cannot be considered ground-truth. Despite these limitations, the model is able to produce plausible patterns and trends of population change.

As of now the grid-level model is being validated. In this approach, the model is used to reproduce observed built-up expansion and population change between 1975 and 2020 according to GHSL products. The model's accuracy in built-up and population allocation is then compared with other approaches to downscale population change. As a first step in this validation procedure, results of built-up development simulated through the current chain of assumptions are being compared with results from other statistical approaches including beta regressions as well as cubist submodels. For specific geographies (the EU, Japan and Korea), census grids are available for multiple points in time. These census grids are used to verify how well the model stacks up to ground truth, rather than dasymetrically mapped population changes. The results of validation will be published in a separate follow-up paper.

1. Introduction

This report details the CRISP (Cities and Rural Integrated Spatial Projections) model. This model is developed with the primary objective of estimating population by degree of urbanisation from 2020 to 2100 for the 2025 World Urbanisation Prospects (WUP). The secondary objective is to develop a flexible and efficient model that can integrate other population projections, for instance, the global Shared Socio-economic Pathways (SSPs) scenarios. CRISP projections are spatially explicit and systematically disaggregated onto 1 km^2 gridded data. They are created by first modelling the development of built-up surface. The modelled built-up surface is then used as an auxiliary variable to downscale population changes. This work is implemented in a dedicated geographic information system (GIS) platform that allows users to create, recreate, and adapt projections.

The modelling sequence is based on exogenous national population projections from the UN World Population Prospects (WPP; UN, 2024). As a rule, the model results are constrained to reproduce those national population projections. Especially in large countries, grid-level models may be unable to capture higher-order trends that cause regional differences in population change within a country. The national-level population projections are therefore downscaled to the level of so-called Functional Areas (FAs), which are large subnational units that are created for the model specifically. These units are designed to represent regions with significant functional autonomy. They are created using a method to define Functional Rural Areas (Dijkstra and Jacobs-Crisioni, 2023), but with more generous parameter settings, so that all land around the globe is divided into roughly 1000 units.

National population projections are downscaled to the FA level through a model that initially extrapolates FA-specific population trends between 2000 and 2020, converging to a single national population trend in 100 years. Built-up surface estimates are derived from the imputed population projections. These estimates of additional built-up surface are estimated through FA-specific built-up surface per capita trends, so that demand for additional built-up may arise because of population growth and/or more per-capita consumption of built-up surface.

The CRISP model downscales FA-specific projections on population and built-up surface to the grid level. This is done for each decade in the projection period. The model first establishes locational suitabilities based on physical characteristics, distance to water bodies and roads, the degree of urbanisation classification, and built-up surface allocation results from prior iterations. The model then allocates additional built-up surface, subsequently population change, and finally assesses the resulting degree of urbanisation classification. The results from this model iteration then inform the model for the next decade. Reference data for allocation are Global Human Settlement Layer (GHSL, R2023) built-up surface estimates (Pesaresi and Politis, 2023) and slightly modified GHS-POP (R2023) population estimates (Schiavina *et al.*, 2023). The remainder of this introduction lays out the model and its assumptions in broad strokes. The following sections describe the modelling procedures in detail.

1.1. Subnational population and built-up surface assumptions

National population numbers are consistent with WPP estimates for every point in time. Population is downscaled to the FA level using the FA's observed population change between 2000 and 2020. Thus, FAs that had higher population growth in a country (or relatively little population decline) obtain an increasing part of the population expected by WPP projections. Initially, the FA's 2000-2020 population change is extrapolated. Regional differences in population change are gradually reduced with the assumption that regions are fully converged by 2130.

Built-up surface estimates per FA are computed as the product of downscaled population size and estimated built-up per capita. Built-up per capita is estimated using a function that describes the regional change in built-up per capita by the speed of population change, the amount of suitable land, and the prior built-up per capita in the FA. After computing future built-up estimates, two conditions are imposed on the results:

- In line with GHSL assumptions, built-up surface in a grid cell cannot decline; and
- built-up surface in an FA cannot exceed available space in the FA.

1.2. Built-up development allocation main assumptions

Built-up surface development projections are downscaled using functions based on empirically observed built-up surface and built-up development. Here, built-up surface represents the fraction of land surface covered by buildings; building heights are not taken into account. To compute a starting state for the model's reference year (2020), this fraction is achieved by dividing total built-up surface in m^2 (Pesaresi and Politis, 2023) by the geographic size (in m^2) of land in the grid cell. Built-up development represents the additional built-up surface that is added to grid cells in a model iteration. It is usually represented as a percentage value.

The approach to model built-up development is based on four premises:

1. mirroring GHSL assumptions (Pesaresi *et al.*, 2024), built-up surfaces cannot be reduced, so that only additional built-up development is distributed;
2. locations differ in suitability, and more suitable locations are more likely to receive additional built-up development;
3. the GHSL built-up surface data shows that a share of 60% of built-up area is a realistic maximum for a 1 km grid cell;
4. over time, the statistical distribution of built-up surfaces shares is remarkably resilient, maintaining similarity to a power-law distribution even in territories with considerable built-up development.

The built-up development modelling is done by a downscaling of additional built-up surfaces. This addition is repeated in each 10-year time step. The downscaling to the grid level is based on an auxiliary ('proxy') variable that includes the locational suitability of every grid cell, a balancing factor, and the expected additional built-up development given the prior built-up surface in the grid cell. Locational suitability is described as a logit probability that is estimated through a 'calibration' process (described by van der Wielen and Koomen, 2024). Expected additional built-up development is described through ET-functions ('Expected Top-ups') that link current built-up surface and average additional built-up development; these are estimated from GHSL built-up developments between 1990 and 2020.

1.3. Population allocation main assumptions

Population is defined as a discrete number of people, without discerning sex or age. Population changes are modelled with universal rules across the entire territory. Global ground-truth population data are elusive, so a statistical model of grid-level population change is unobtainable. By necessity, the implemented rules, therefore, reflect coarse assumptions on how population relates to locational suitabilities and responds to the additionally allocated built-up development. Expected exogenous

population changes are thus downscaled using modelled built-up changes as a support variable. In addition to exogenous population changes, a parametrised share of the region's population is re-allocated by the model to capture internal migration. The population model consists of three steps:

1. first, a population pool is formed that represents all people who are allocated in one iteration in an FA; then,
2. part of the population from that pool is allocated to new built-up development; and finally,
3. the remainder is distributed while taking into account locational suitabilities and population change limits.

The latter population change limits indicate maximum decennial population changes observed in the GHSL population layers, depending on whether a grid cell already contains population, and whether the grid cell is the most populous in its environment. Further sections in this report describe the model in more detail.

1.4. How to use the model

The model is based on open data and the open-source GeoDMS platform (ObjectVision, 2023). The modelling process requires up to 120Gb of memory for large continents. System requirements are a recent Windows machine with at least 64Gb RAM (or a very large page file). For development purposes, the model is typically run by dividing the earth into six separate regions (named 'continents') with fixed parameters, but the application allows for substantial modification and customisation. A tutorial is available to get new users started with the model; it is accessible through https://github.com/ObjectVision/2BURP/wiki/Tutorial. For further guidance while using the model, enthusiastic users are encouraged to get in touch with the authors of this report.

2. Geography is treated explicitly

Geography is treated explicitly in the CRISP model. All model outputs are mapped on grid cells that each describe 1 km^2 of the Earth's surface. In the reference application, geography is described in such a way that all grid cells represent an equal area. This is done using the so-called Mollweide geographic projection system[1]. For ease of use, all model results can later be converted to the so-called WGS84 projection system, with coordinates based on latitudes and longitudes, and variation in grid cell size depending on distance to the equator. The CRISP model architecture, in principle, also allows modelling scenarios in the WGS84 projection system directly.

Three geographic definitions are specific to the CRISP model and its reference application. These are the selection of grid cells included in the modelling, the used Functional Areas (FAs) by which exogenous subnational projections are constrained, and the internal boundaries used to demark countries and continents. The following sections describe these geographical units in detail.

2.1. Grid cells selected for modelling

The model derives its observations and most data from a global raster in which the entire globe is divided into equally sized units ('grid cells'). This raster is referred to internally as the 'domain'. However, for the sake of computational simplicity, the model does not include all grid cells within a study area. Grid cells are *excluded* if they

- are not within the model's territorial boundaries, so that e.g. Antarctica and some uninhabited islands are excluded;
- are permanently covered by surface water or ice; or
- are more than 5 km away from a grid cell that has built-up surface in 2020.

The resulting selection of grid cells that are included in the modelling is referenced internally as the 'compacted domain'.

The boundaries-based criterion ensures that grid cells are not included if they are not covered in the exogenous projections for built-up surface and population. The source of all country borders is the GPW 4.11 database (CIESIN, 2018)[2]. However, these borders have been adapted by JRC's GHSL team, essentially moving country boundaries 3 km into large waterbodies where possible. This has the advantage that, e.g., outcrop raster cells are most likely included within country borders, thus minimizing the amount of land that falls outside of coastlines due to misalignment and land reclamations.

We also exclude grid cells that are permanently covered by water and ice, as built-up development is highly unlikely in such locations. This does imply that we are not taking into account, e.g., efforts to reclaim land from water.

[1] To be precise, the projection system with EPSG code 54009, an equal area projection with a very low average distortion of shapes.

[2] Thus borders are adopted from a well-established source. To keep modelling practical, ongoing border conflicts are ignored here. The adopted border lines should not be construed as an opinion regarding such border conflicts.

Finally, we exclude all grid cells that are more than 5 km away from any known built-up grid cells in 2020 according to GHSL built-up surface grids. This criterion is implemented to restrict the number of observations used for modelling in land areas with very limited human occupation, such as the Sahara Desert. The threshold is set to 5 km based on an analysis of historical built-up development as observed in GHSL built-up surface data. In this analysis, incremental 1 km buffers were drawn around grid cells that had non-zero built-up surface in 1975, with the largest buffer describing grid cells that are 49 to 50 kilometres away from a location with some pre-existing built-up surface. Subsequently, the amount of built-up development between 1975 and 2020 was summed per incremental 1 km ring (see Table 1).

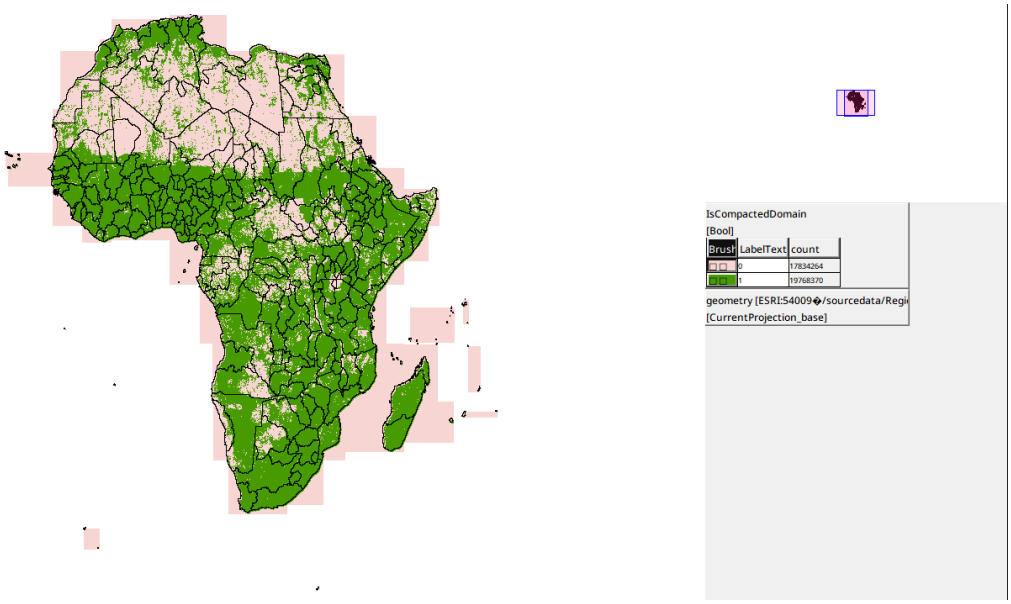
Table 1. Built-up surface development between 1975 and 2020 per 1 km ring around grid cells with pre-existing built-up surface. Source: elaboration of GHSL built-up grids, release 2023.

Distance from grid cell with some built-up surface in 1975 (km)	*Built-up surface increase 1975-2020 (km^2)*	*Built-up surface increase 1975-2020 not included in buffer (%)*
1	1,427,368	2.01%
2	19,957	0.64%
3	4,775	0.31%
4	2,320	0.15%
5	681	0.10%
6+	1,482	0.07%

Source: own calculations based on GHS-BUILT-s R2023A

This analysis shows that, in the 45 years between 1975 and 2020, only a negligible amount (0.07%) of built-up development occurred more than 5 km away from prior built-up surface. We therefore believe it plausible that a similar 5 km buffer around grid cells with some built-up surface in 2020 is sufficient to describe the space for built-up development between 2020 and 2100. Excluding grid cells outside of this 5 km buffer increases model speed considerably. The dropped grid cells are primarily in highly remote, uninhabited places. As an example, the grid cells included in the compacted domain (1) and the unselected cells in the tiled raster domain (0) are shown for Africa in Figure 1.

Figure 1. Result of selecting the compacted domain in Africa, based on the criteria outlined in this section. Dark green grid cells are selected; light pink grid cells are not.



Source: CRISP model output, JRC analysis.

This selection procedure does have one caveat. In specific geographies, the subnational units used as inputs in the GHS-POP grids have population, but no observed built-up surface. In these cases, population is distributed proportionally over sometimes vast administrative units used in the GHSL dasymetric mapping approach (Freire *et al.*, 2016). As there is no built-up surface to satisfy the nearby built-up criterion, these proportionally distributed people are excluded from the reference population grids used in the modelling. In cascade, the model yields slight discrepancies in historical population grids.

2.2. Functional areas (FAs) as subnational units

The model defines territorial boundaries based on large spatial units within country borders, referred to as functional areas (FAs). These units were designed to create regions that may have a different demography trajectory, and are a key component of the model's framework. FAs are created using a method initially developed for defining Functional Rural Areas (Dijkstra and Jacobs-Crisioni, 2023), but with more generous parameter settings, dividing all land around the globe into roughly 1000 units. These functional areas centre around urban regional centres. They were created using OpenStreetMap network data (OSM, 2023). First, the cities and towns with the largest population within a 30-minute travel-time by car are defined as a regional centre (Jacobs-Crisioni, Kompil and Dijkstra, 2023). Catchment areas are subsequently generated for all those regional centres. Those catchment areas are restricted by national boundaries.

Subsequently the generated catchment areas are combined to form functional areas. We refer to those growing combinations as 'catchment sets'. Combination happens through an iterative process in which, repeatedly, two catchment sets are paired together. In every iteration, a pair of catchment sets is chosen from all possible combinations, if they have the shortest travel time between their centres, and if they meet specific target population and drive time criteria. All other possible catchment combinations remain unchanged in the round, and are considered for combination again

in the next round. By repeating this process, catchment sets are added together until 1) all catchment sets are more than the minimum driving time apart and 2) all catchment sets have at least the minimum population target or have no other catchment set within the maximum drive time threshold.

Through variations of the population and drive time thresholds, four definitions of functional areas are generated. Per country, a specific definition was chosen, partially to limit the number of subdivisions in very populous countries, and partially to ensure that geographically very large countries with limited population are divided into multiple subnational units. Thus, the end result consists of functional areas with definitions varying per country.

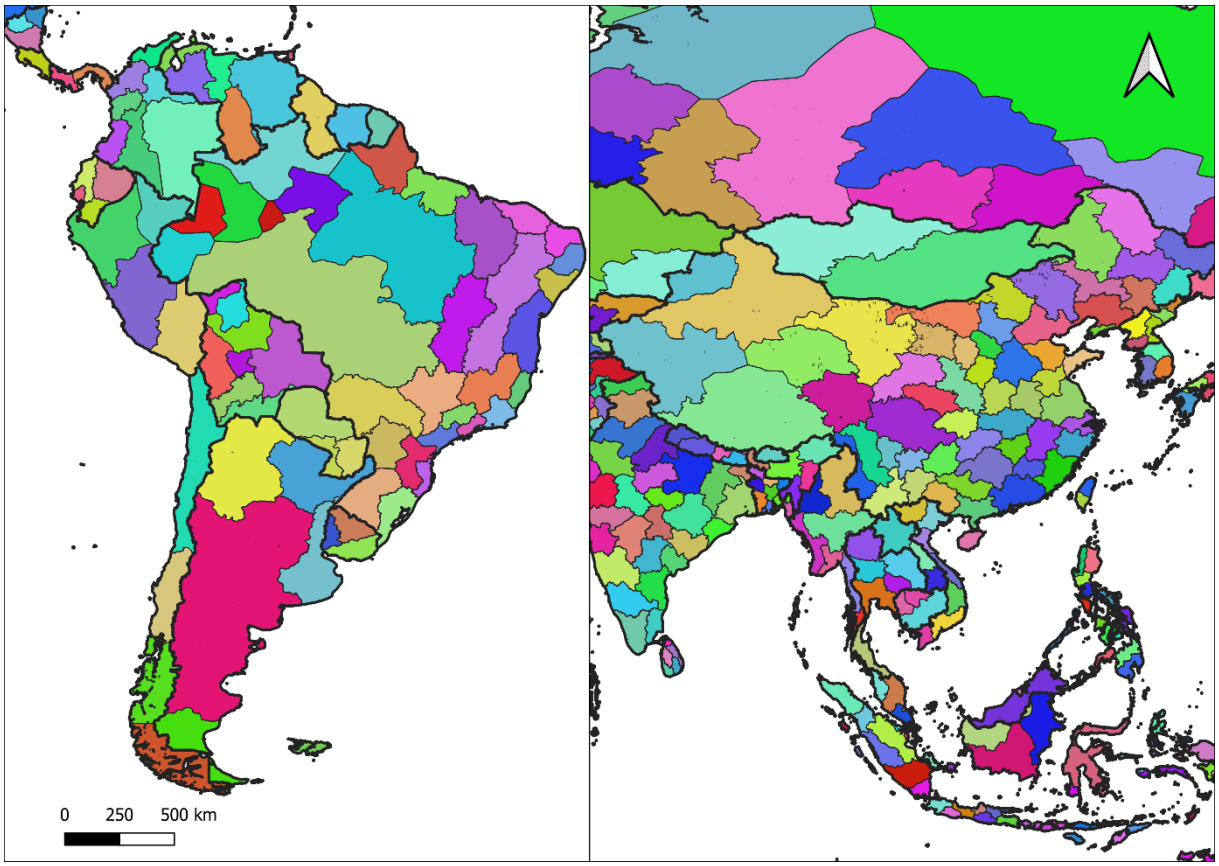
The following criteria were used and selected:

- 2M inhabitants target population, 120 minutes minimum, 240 minutes maximum driving time (used for 163 countries).
- 4M inhabitants target population, 120 minutes minimum, 480 minutes maximum driving time (used for 54 countries, namely MAR, AFG, AGO, ARG, ARUS, AUS, BGD, CAN, CHL, CIV, CMR, COD, COL, DEU, DZA, EGY, ERUS, ESP, ETH, FRA, GBR, GHA, IDN, IRN, IRQ, ITA, JPN, KAZ, KEN, KOR, MDG, MEX, MMR, MOZ, MYS, NGA, NOR, PAK, PER, PHL, POL, SAU, SDN, THA, TUR, TZA, UGA, UKR, USA, UZB, VEN, VNM, YEM, ZAF).
- 4M inhabitants target population, 120 minutes minimum, 240 minutes maximum driving time (used for 17 countries, namely BLR, BOL, MLI, MRT, NER, NPL, NZL, PNG, PRK, ROU, SEN, SOM, SSD, SWE, TCD, TKM, ZMB).
- 4M inhabitants target population, 240 minutes minimum, 480 minutes maximum driving time (used for 3 countries, namely BRA, CHN, IND).

Five small countries with less than 2M inhabitants consisted of multiple, relatively small functional areas. These countries were COM, CPV, GNQ, LVA and ESH. The functional areas in these countries were grouped into single functional areas per country.

Finally, the functional area algorithm occasionally yielded erratic geographies and implausible exclaves, typically in low-density areas with capricious shorelines, islands and navigable rivers. Where the algorithm created such distortions, functional area boundaries have been adjusted manually. This has been done in the Amazon estuary in Brazil and the border of MEX and GTM. Similar adjustments have been made to the surroundings of Ouésso, Congo; the southern tip of the CAR; the northeastern border of ZAF with eSwatini; areas in SEN; MRT; MAR. This has also affected areas in Bangladesh, India, Pakistan, Iran, Azerbaijan, China and Russia, Spain, Germany, Norway, the UK and Czech Republic. In addition, a number of very small functional areas in Bangladesh and India have been merged with larger surrounding functional areas. Examples of functional areas are shown graphically in Figure 2.

Figure 2. Functional area examples showing South America (left) and parths of Southern and Eastern Asia (right). Country borders are shown by broad black lines, individual FAs are shown by varying colours and bounded by narrow lines.



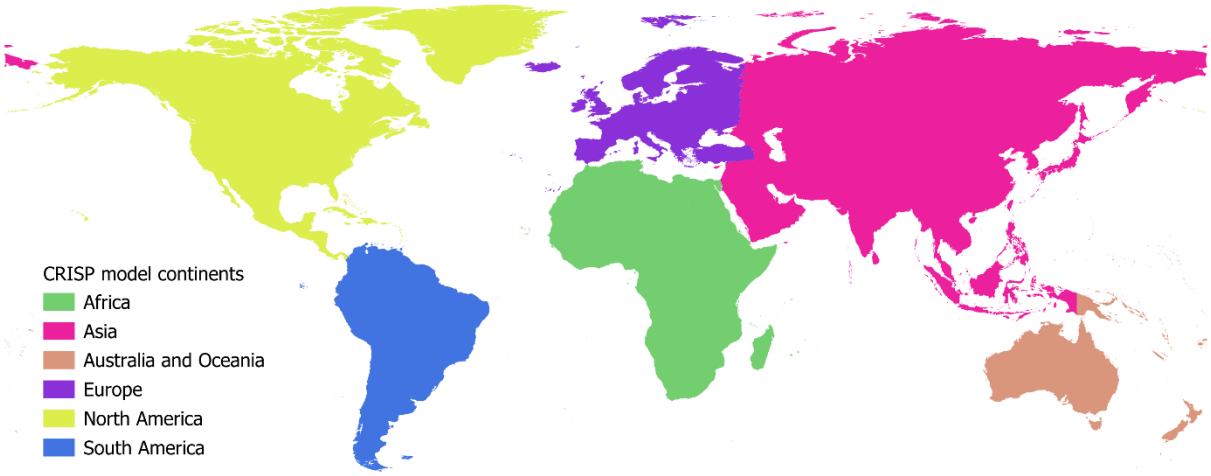
Source: CRISP model input, JRC analysis.

2.3. Internal boundaries for computational purposes

The model uses a set of internal boundaries that represent continents, countries, and the subnational FA units described before. These internal boundaries represent the best set of boundaries available for our modelling purposes, but do not represent a position on existing border conflicts or contested areas. All internal boundaries derive from the FAs described in the previous section. Continental boundaries were used to separate training sets for model calibration, and many input parameters in CRISP are defined per discerned continent. By default, the model provides projections for all countries in an identified continent. Alternatively, country borders can be used to limit model runs, for instance while testing model specifications.

The country and continent definitions that are used internally are recreated by combining functional area geometries into larger landmasses. Many of the choices in the model's internal boundary system thus cascade from choices made in the functional area creation process. The model inherited this separation. The most notable result is that, in line with the OpenStreetMap data used to make the FAs, the Russian Federation is split into two separate zones, which are in the Asian and European continents. For the same pragmatic reason, the entire country of Turkey is considered European. In this study, no interactions are assumed between FAs. We therefore do not expect that our pragmatic separation of geographies will have unwanted effects on the modelling results. Figure 3 shows the current representation of continents in the model, highlighting Europe.

Figure 3. Continents as they are currently defined in the CRISP model.



Source: CRISP model input, JRC analysis.

3. Population and built-up are derived from national population projections and allocated at the subnational level

National population projections for 2020-2100 are derived from the UN World Population Prospects (UN, 2024). In the case of multiple FAs in a country, national projections are downscaled over FAs using, as a proxy, a function that initially extrapolates FA population changes from historical FA population change rates, converging to national growth rates in the long run. This is expressed as the expected population $\widehat{P_{t+1,f}}$ at t+1 in functional area *f* (1),

$$\widehat{P_{t+1,f}} = P_{t,f} e^{\widehat{grFA_{t,f}}}, \quad (1)$$

which in turn depends on the expected population growth factor in $\widehat{grFA_{t,f}}$ (2):

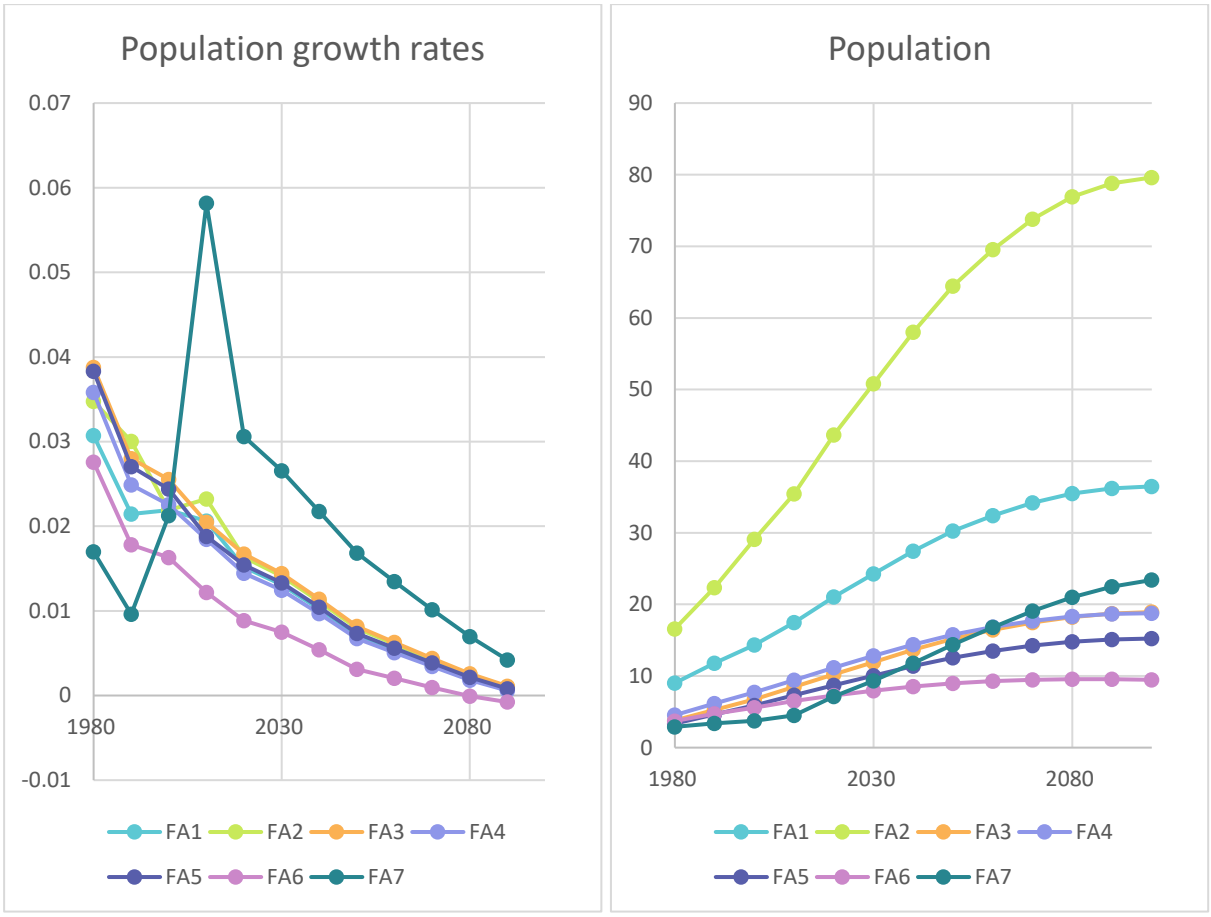
$$\widehat{grFA_{t,f}} = (1 - w_t)\left(-0.07 \leq \frac{\ln(P_{t,f}/P_{t-10,f})}{10} \leq 0.04\right) + w_t grWPP_{t+1,N}. \quad (2)$$

This growth factor depends on:

- Observed annual population change in the FA between 2000 and 2010, and between 2010 and 2020, of which the natural log is capped between - 0.07 and +0.04, so that population change in a decade cannot exceed +/-50%. Historical population is aggregated from GHSL population grids.
- Projected population change at the country level (*N*), as given by the World Population Prospects (UN, 2024).
- A weighting factor *w* that ensures a linear convergence of functional areas to the national level in 100 years, so that $w_{2019} = 0$ and increases linearly until $w_{2119} = 1$.

The result of $\widehat{P_{t+1,f}}$ is then used to downscale the national population according to the WPP projections. By means of example, the observed and modelled population growth rates and population totals are shown graphically for the case of Egypt in Figure 4.

Figure 4. Observed and modelled population growth rates (left) and total population (right, in millions) in Egypt's functional areas.



Source: own elaborations of GHS-POP trends and WPP (2024) , JRC analysis.

Downscaled population projections are subsequently used to estimate future built-up surface in an FA. The consumption of built-up surface per capita is used to link the two. The currently used GHSL built-up surface data distinguishes between residential and non-residential functional use of the area that is built upon. In the projections for the WUP 2025, only residential built-up surface estimates are added to the area of the total (residential and non-residential) built-up surface observed for the year 2020, by keeping constant the non-residential component as observed in 2020 GHSL data ($BUnres_{2020}$).

The total amount of built-up surface in an FA, $BU_{f,t+1}$, is subsequently projected through the evolution of residential built-up surface per capita ($BUCAP_{f,t+1}$) in the FA, and the population dynamic from ($P_{f,t+1}$) (3):

$$BU_{f,t+1} = P_{f,t+1} * BUCAP_{f,t+1} + BUnres_{2020}, \quad (3.1)$$

$$BUCAP_{f,t+1} = BUCAP_{f,t} * \left(1 + grBUCAP_{f,t+1}\right) \quad (3.2)$$

$$BUCAP_{f,t} = BUres_{f,t}/P_{f,t}, \quad (3.3)$$

in which $P_{f,t}$ is the population and $BUCAP_{f,t}$ is the residential built-up surface per capita, both of functional area *f* in the decade *t*, and $grBUCAP_{f,t}$ is the estimated compound decennial growth rate of BUCAP between decade *t* and *t+1*. This decennial growth rate is obtained using the following polynomial regression, based on historical dynamics (2010-2020) of large units (population ≥ 250,000) (4):

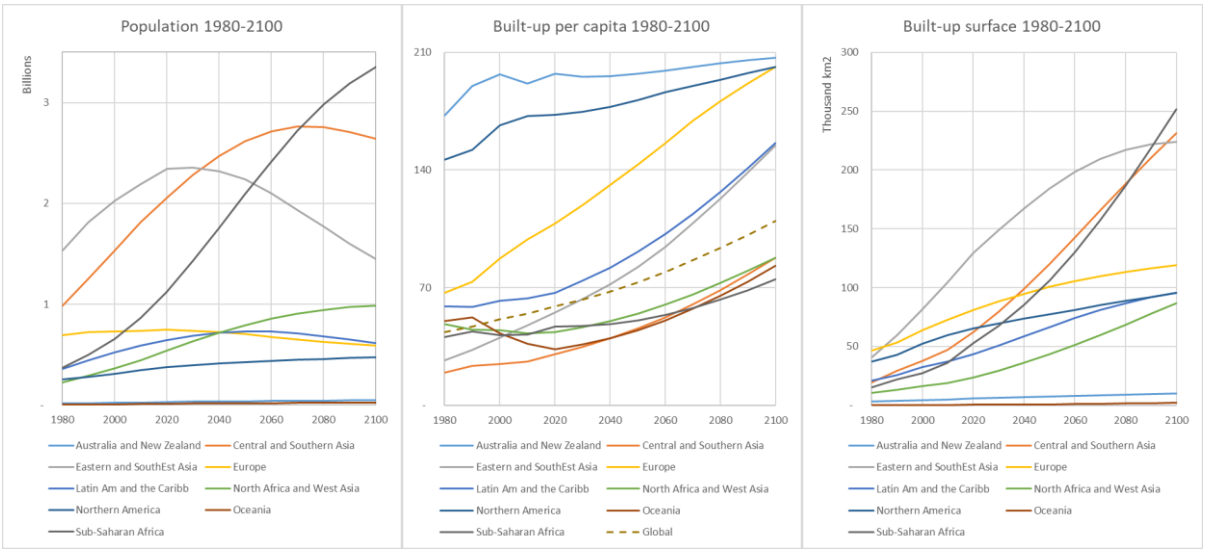
$$grBUCAP_{f,t+1} \cong f(\ln BUCAP_{f,t} + \left[\left(P_{f,t+1}/P_{f,t}\right)^{\left(\frac{1}{10}\right)} - 1\right] + \ln\left[BU_{f,t}/MA_f\right]), \quad (4)$$

which depends on residential built-up surface per capita of the functional area *f* in the decade *t*, the population dynamic between the decade t and decade *t+1* (as the compound annual growth rate of the decade), and the density of residential built-up surface over the FA modelled area (MA_f). The *MA* variable indicates the suitable area in the FA that is in the so-called compacted domain, and thus available for built-up development. The suitable land is obtained from the Gross Land Development Potential, a derived layer of the GHSL built-up surface production[3]. The parameters used in this function are derived from repeated bootstrapped regressions (500 iterations based on samples with replacement) on historical changes in residential built-up surface per capita, and are significant at least at the 0.05 level.

The implemented method yields a stabilisation of built-up surface per capita in Northern America, Australia and New Zealand, and considerable increases in built-up surface per capita in most other world regions (see Figure 5). Multiplied by population growth, the model yields considerable growth in built-up surface, especially in Sub-Saharan Africa and large parts of Asia.

[3] Land development potential is estimated using a symbolic machine learning model (Pesaresi *et al.*, 2024) explaining the presence of built-up surface by elevation (ESA, 2024), slope (derived from ESA, 2024) and permanent snow cover (EU-CLMSI, 2015).

Figure 5. Examples of observed and projected population (left), built-up surface per capita (middle) and built-up surface (right) for the world regions included in the CRISP model.



Source: CRISP model inputs, JRC analysis.

4. Built-up surface change is allocated using empirical support

In every iteration, the model allocates built-up development to grid cells. The affected grid cells may already contain built-up surface. The total area of built-up surface that is added in a decade at the FA level is assumed to be known beforehand, and needs to be provided by the user, for example, through the procedure discussed in Section 3.

The model allocates regional built-up developments to the most likely locations, while maintaining the modelled built-up surface distributions similar to the 2020 distribution. To maintain similarity to the current value distribution and account for the expected increase in built-up surface, the model setup combines estimated locational suitabilities, so-called *ET* values, and a counterbalancing mechanism. As a result, the additional built-up development is disaggregated using, as an auxiliary variable, the proxy of local built-up change $\widehat{\Delta BU}$, which varies per grid cell in *i.* This proxy is computed as (5):

$$\widehat{\Delta BU}_i = S_i * \left(S_i / \bar{s}_f\right) * ET_f, \tag{5}$$

containing:

- The logit locational suitability of every grid cell in *S*, computed by adding empirically derived variable impacts. The underlying logit function describes the probability that a grid cell is at least 2.5% built-up.
- A balancing operation, in which grid-cell locational suitability is divided by the average locational suitabilities ($\bar{s}_f$) for all grid cells with the same prior built-up level in the FA. These built-up levels are classes of built-up surface with a 1% interval. This balancing operation equalises probabilities between prior built-up levels. This is needed because grid cells with more built-up surface tend to have structurally higher locational suitabilities, causing the surface distribution to bulge in the tail after a number of model iterations.
- The so-called ET ('Expected Top-up') value, which depends on the prior built-up level of the grid cell. ETs contain the average decennial built-up development per 1% built-up level, calculated separately per continent. They tend to increase rapidly from 0%, reach their highest values between 10% and 30%, and then slowly drop to 0 between 60% and 65% built-up.

4.1. Locational suitability

The percentage locational suitability S_i for the presence of built-up surface is determined using a logistic regression analysis. In this study, this is referred to as the calibration. The approach has been adopted from the work on the 2UP model that preceded the development of CRISP (Andree and Koomen, 2017). It thus leans on the AutoGLM methodology, a bespoke method for the selection and estimation of balanced logistic regression models. We use this method to explain the occurrence of built-up area in a single year, which corresponds to the initial year for simulation. Ferdinand et al. (2021) provide a brief motivation for this so-called static calibration approach and describes its characteristics and application to explain urban development patterns. A more in-depth motivation behind this choice, as well as more detailing of the model's inner workings, is provided by Ferdinand, Andree and Koomen (2021).

4.1.1. Dependent variable

The dependent variable we use in the suitability assessment measures the share of built-up land in 2020 relative to a cell's total area and classifies these in a similar, binary manner as the calibrations for 2UP. This classification distinguishes between *developed* and *non-developed* land, where cells that exceed a pre-specified critical share of built-up land are assigned to the developed class, and cells that do not exceed this critical share are assigned to the non-developed class.

As CRISP aims to simulate all significant occurrences of built-up area, and not a smaller subset of locations that are deemed urban as in the 2UP model, there is a need to adapt the threshold. This is especially relevant as the latest GHSL release offers a more refined account of built-up area per cell (Pesaresi and Politis, 2023). Even in the centres of larger urban agglomerations the built-up area fraction now rarely exceeds the 50% threshold that was considered a condition for urban land in 2UP, and therefore was applied in the previous calibration efforts.

We therefore opt for a low, 2.5%, threshold. This allows us to recognise most of the cells that contain at least some built-up area. This threshold strikes a balance between correctly identifying small clusters of reasonably developed built-up land surrounded by large undeveloped areas, whilst simultaneously excluding undeveloped areas of nature or agriculture that contain the occasional barn, farmhouse, or cabin.

4.1.2. Explanatory variables

The selection of explanatory variables included in this calibration builds upon the drivers used in the most recent calibration of the 2UP model (Ferdinand, Andree and Koomen, 2021). In general, these drivers can be categorised into broad classes such as local population density, distances to relevant spatial features (coastline, roads, cities, etc.), topographical characteristics (elevation, slope, terrain roughness index), prevalence of specific natural hazards (floods, earthquakes, etc.), and planning policies (protected areas). In our analysis we use up to fifteen distinct drivers to describe the presence of developed areas. The exact number of included explanatory variables differs per continent and depends on their collinearity and significance in explaining developed area presence in that world region. The used variables and their definition are given in Table 1.

Most of the included variables remain functionally unchanged with respect to their 2UP counterparts, but whenever possible we updated, improved or expanded the spatial data reflecting these drivers. Distance to freshwater, which previously measured Euclidean distance to either the nearest river or nearest lake, has been replaced by distance to large inland water. This new factor harmonises these separate definitions in addition to including any inland water that may not have been classified as a river or lake. The variables describing topographical characteristics now make use of a global elevation dataset with a finer resolution. Landslide probabilities were changed to correspond with the landslide risk reported in the Global Landslide hazard map. A more extensive description, with sources for each dataset used, is included in Annex II and van der Wielen and Koomen (2024).

Table 2. Characterisation of the explanatory variables used in the locational suitability estimation.

Variables	Definition	Unit
Intercept	Constant	n/a
ln of neighbourhood pop. density	Continuous	Ln (People)
Distance to coast	Continuous	Kilometres (≤ 250)
Distance to large inland water	Continuous	Kilometres (≤ 250)
Distance to major road	Continuous	Kilometres (≤ 250)
Distance to secondary road	Continuous	Kilometres (≤ 250)
Grid-cost distance to nearest city	Continuous	Abstract cost (≤ 100)
Grid-cost distance to nearest city/town	Continuous	Abstract cost (≤ 100)
Grid-cost dist. to nearest city/town/village	Continuous	Abstract cost (≤ 100)
Elevation	Continuous	Metres
Slope	Continuous	Degrees
Terrain roughness index	Ordinal	TRI (1-7)
Protected area	0/1	No / Yes
Flood-prone area	0/1	No / Yes
Earthquake intensity	Ordinal	Mercalli (0-12)
Landslide prone	Ordinal	ThinkHazard (1-4)

Source: CRISP model setup.

Newly added in the calibration effort, is a novel measurement of travel distance in the form of a grid-cost distance. This characterises the effort to travel between an origin and destination cell, where different routes incur costs based on the characteristics of the cells they pass. Travelling over a cell with terrain that is more difficult to traverse (i.e. requiring to cross water, change elevation, and so on) will incur a higher overall cost for the route. This is based on a grid of abstract impedances or travel costs *C*, of which a value of 1 is equivalent to a 1km distance traversed on flat ground. Minimising these costs will give an indication of the most efficient route between an origin and destination cell. The precise formulation is inspired by the approach taken in Nelson (2008) and takes this form (6):

$$C = 25BORDER * F^{elev} * F^{slope} * F^{water} \quad (6.1)$$

$$F^{elev} = \begin{cases} 0.15 * e^{0.0007 * elevation} & if\ elevation > 2000\ m \\ 0 & if\ elevation \leq 2000\ m \end{cases} \quad (6.2)$$

$$F^{slope} = 1/e^{-3 * slope\ gradient} \quad (6.3)$$

$$F^{water} = 1 + 3 * fraction\ of\ water \quad (6.4)$$

So, travel costs are 25 times higher if traversing a grid cell means traversing a national border (as indicated by the *BORDER* dummy). Travelling at more than 2000 metres above sea level is penalized by elevation, and travelling at a slope is penalized by the exponent of the gradient of the slope. Finally, the amount of water present in a grid cell leads to penalties on travel cost as well.

The main benefit of this measure over travel time is that it does not rely on external information on transport networks, for which we do not have information on future conditions that may be influenced by, for example, congestion, changes in mobility modes, infrastructure expansion, or other changes. It thus captures the benefits of accessibility at a more basic level that is likely to offer a more robust understanding, which we assume will maintain validity in the more distant future.

The new accessibility indicators generated with the grid-cost approach include reference to settlements of different levels of urbanisation. Rather than only referring to cities of a specific size we now refer to villages, towns and cities following the degree of urbanisation classification (Dijkstra et al., 2021). As the definitions of grid-cost distance and degree of urbanisation rely on static conditions (such as slope or presence of water) and dynamic aspects (built-up area and population) that are simulated endogenously by CRISP, these accessibility indicators can also be calculated for a future year. The methods to evaluate travel costs and assign degrees of urbanisation are approximated by routines internalized in the CRISP model. This element of the suitability equation is updated in every timestep during the simulation process.

4.1.3. Results

We now move on to discussing the estimated coefficients as reported in Table 3. Metrics such as model performance are discussed in van der Wielen and Koomen (2024). The variables included differ between continents as a result of the restrictions we placed and preprocessing steps we put in place to handle variables with high bivariate intercorrelation and low explanatory power. Almost all variables included in the final models are significant at the 0.1% level ($p<0.001$), barring some of the variables for Australia/Oceania and South America, which do not reach such a high significance level.

Table 3. Regression results for logit model with down sampling, explaining presence of built-up grid cells in 2020.

Variables	Africa	Asia	Australia / Oceania	Europe	North America	South America
Intercept	1.759***	0.310***	7.393***	-4.620***	2.761***	1.674***
ln of neighbourhood pop. density	0.706***	0.846***	1.090***	1.754***	0.865***	1.604***
Distance to coast	-0.004***	-0.003***	-0.015**	-0.003***	-0.008***	-0.016***
Distance to large inland water	-0.003***	0.001***	-0.005***	-0.002***	-0.004***	-0.004***
Distance to major road	-0.003***	-0.013***	-0.008***	-0.036***		-0.017***
Distance to secondary road	-0.005***	0.001***	-0.004***	-0.003***	-0.006***	0.008***
Grid-cost to city	-0.009***	-0.011***	-0.020***	-0.004***	-0.010***	-0.016***
Grid-cost to city/town			-0.008***	-0.012***	-0.012***	
Grid-cost to city/town/village	-0.070***	-0.010***	-0.011***	-0.042***	-0.037***	-0.018***
Elevation	0.0004***	-0.0004***	-0.001***	-0.001***	0.0004***	0.0002***
Slope	-0.022***		-0.150***			-0.035***
Terrain roughness index	-1.458***	-1.941***	-2.110***	-0.446***	-1.566***	-1.486***
Protected area	-0.204***	-0.170***			-0.415***	-0.193*
Flood prone area	-0.165***			0.128***	0.667***	
Earthquake intensity	0.084***	0.055***	-0.113***	0.020***	-0.022***	-0.131***
Landslide prone	-0.206***	-0.067***	-1.049***	-0.455***	-0.352***	-0.198**

Significance coding: *** p<0.001, ** p<0.01, * p<0.05.

Source: Van der Wielen and Koomen, 2024, where model performance is discussed and further details are given.

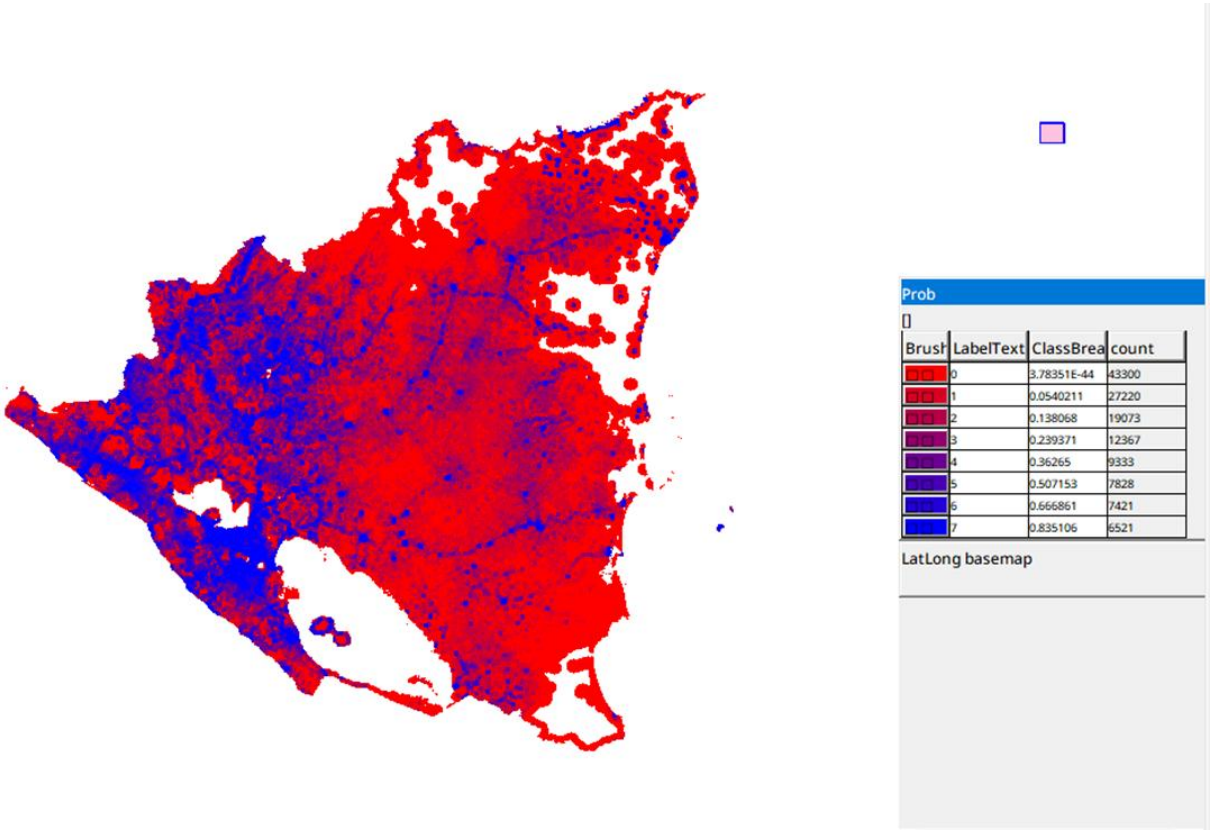
For most variables, the direction of estimated effects is consistent across all continents, though we do notice some discrepancies. For example, the distance effects for large inland water and secondary roads are negative for all continents, except for Asia. We find relatively small coefficients for elevation on all continents, but the direction is flipped for half of the continents. The presence of flood-prone areas is negative for Africa, but positive for Europe and North America. A full discussion of continent-specific variations between the importance of drivers is beyond the scope of this report.

Accounting for this variation allows us to better identify locations suitable for built-up area development on each continent.

4.1.4. Implementation

The resulting coefficients per attributing factor are combined with the geographical data of these factors to generate a map that shows the unscaled locational suitability values of built-up grid cells being present; expressed as the function $L_{i,t} = [\beta 0 + \beta 1X1 + \beta 2X2 + \cdots \beta nXn]$. Subsequently, this is transformed using a logit function to express the percentage locational suitability or probability of the built-up grid cells being present through $S_{i,t} = e^{L_{i,t} \leq 80}/(1 + e^{L_{i,t} \leq 80})$. Here, *L* is capped at 80 arbitrarily to prevent numerical overflows when taking the exponential from its result. The resulting locational suitabilities in Nicaragua are shown in Figure 6 for 2020. Variables that change as a model outcome, such as for instance neighbourhood population densities and the grid cost to cities, towns and villages, are updated in every model iteration. Thus, the locational suitabilities are recomputed in every model iteration.

Figure 6. Exemplary map of values of locational suitabilities describing Nicaragua in 2020.



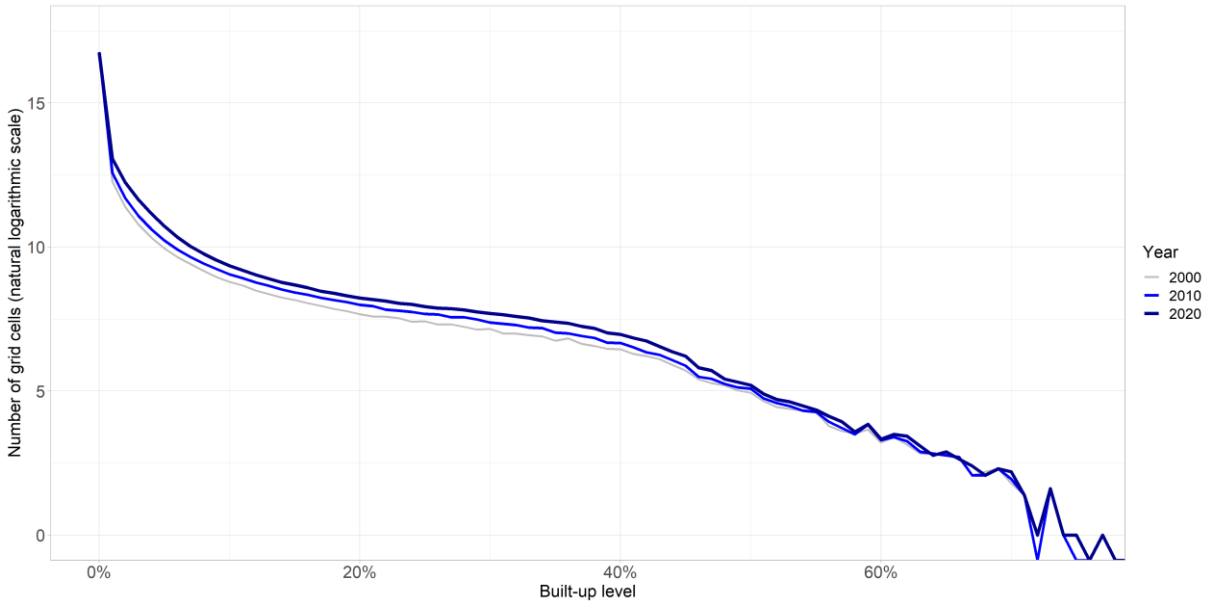
Source: CRISP model output, JRC analysis.

4.2. Locational suitability balancing factor

As noted before, the distribution of built-up values is remarkably resilient and somewhat resembles a power-law distribution. To establish this, built-up values derived from the GHS-BU data have been binned into 1%-wide classes in 2000, 2010 and 2020, henceforth referred to as built-up levels, and

subsequently plotted for the six continents accounted for. The statistical distribution of values can be seen for Africa in Figure 7. The other five continents have very similar patterns over time.

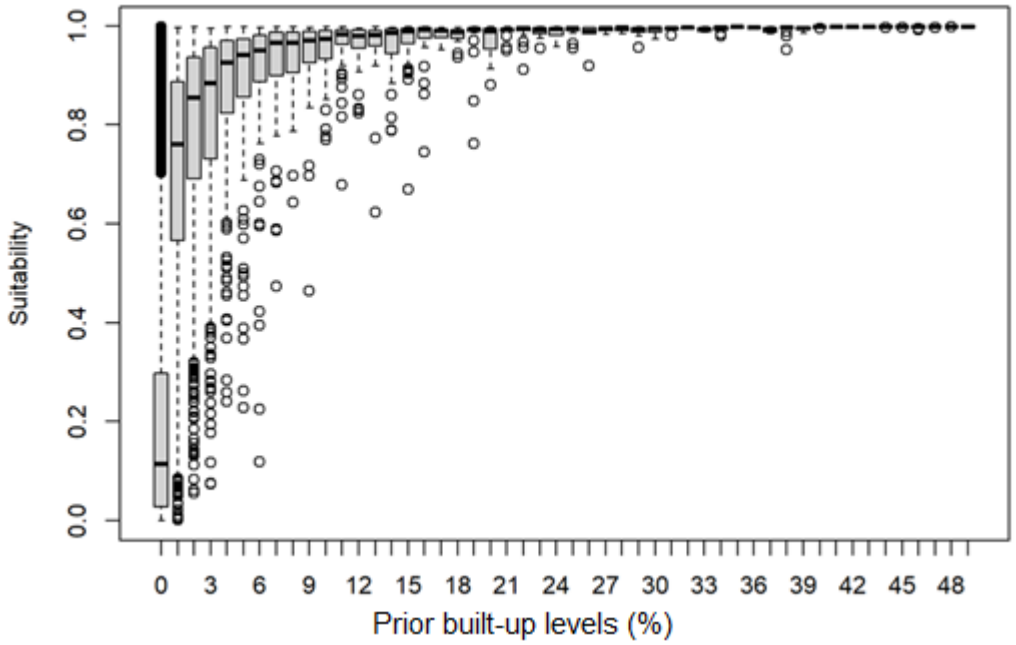
Figure 7. Number of observations by 1% built-up level in Africa in 2000, 2010 and 2020 according to GHS-BU data, showing remarkably similar distributions despite considerable built-up development.



Source: own elaborations of GHS-BUILT-S R2023A, JRC analysis.

The modelling framework attempts to reflect the resilience in distributions through two somewhat unorthodox steps. The first step is to normalise suitabilities by built-up level. This is necessary because suitability values are correlated with pre-existing built-up levels. Thus, grid cells in the low built-up levels tend to have structurally very low locational suitabilities. Figure 8 shows the distribution of locational suitability values, per built-up levels, for Nicaragua.

Figure 8. Locational suitability distribution by prior built-up levels in Nicaragua.

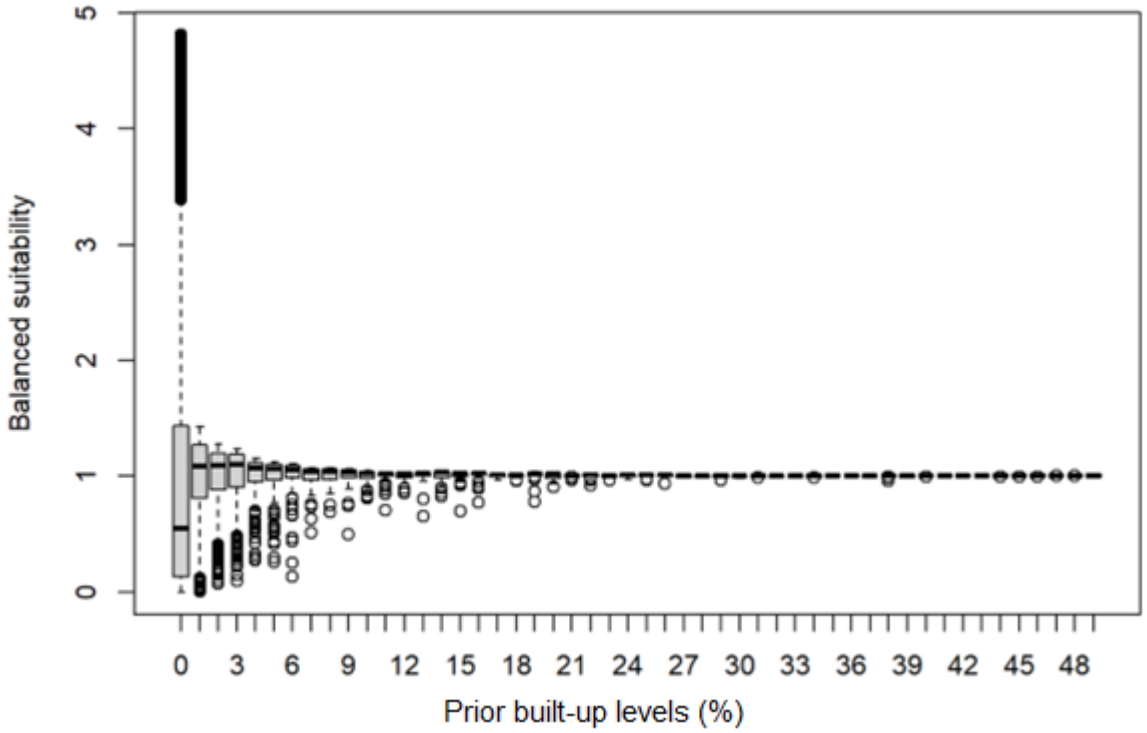


Source: CRISP model output, JRC analysis.

The skewed locational suitability distribution implies that grid cells with low prior built-up values would receive very little additional built-up development, in case the model would simply downscale built-up surfaces using unmodified locational suitability levels as a proxy. This would shift the built-up value distribution to higher built-up levels. Ultimately, it would somewhat choke the newly built-up development, as the supply of the built-up values assigned to the middle built-up levels dries up.

A counterbalancing mechanism has therefore been introduced to account for structural variation in locational suitability at each pre-existing built-up level. This mechanism uses the multiplicative inverse or reciprocal of the average locational suitability per built-up level to offset the correlation between locational suitabilities and existing densities. The result is a balanced suitability. Practically, this operation tends to give a bump to locational suitabilities assigned to grid cells with low prior built-up levels (see Figure 9).

Figure 9. Balanced distributions of locational suitability by prior built-up levels in Nicaragua.



Source: CRISP model output, JRC analysis.

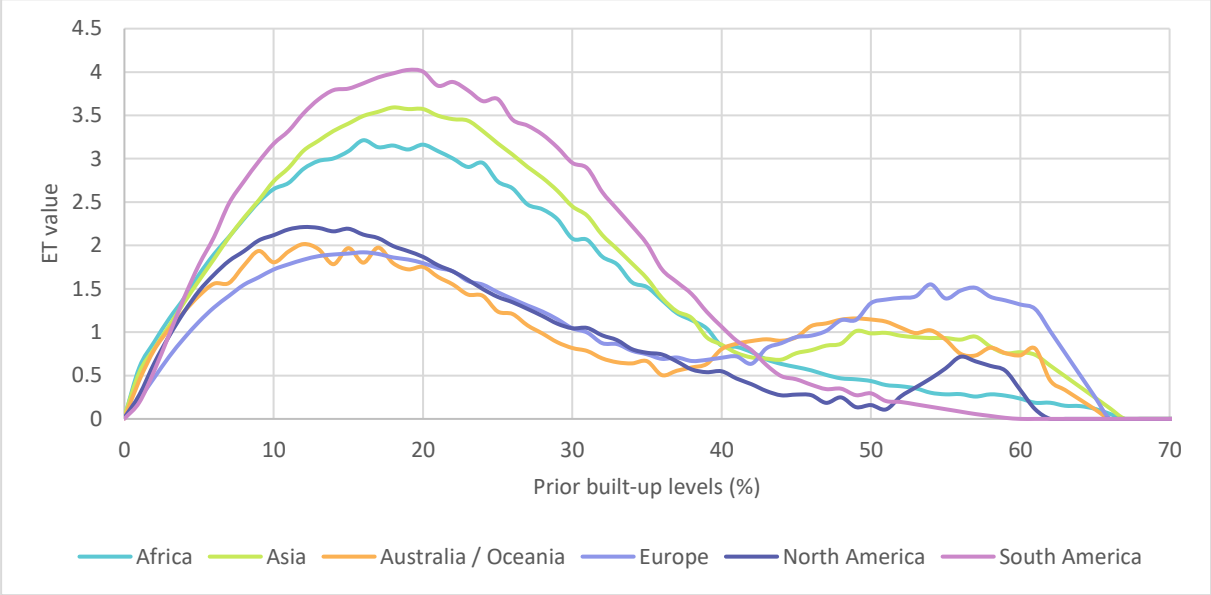
4.3. Expected Top-ups

Prior built-up surface affects subsequent built-up development. For instance, at very low built-up levels, the critical infrastructure needs to be improved, which in turn generates increased costs. At high built-up levels, new infrastructure developments are restricted in terms of available space (and thus degrees of freedom in construction design) and, for example, because of NIMBYism (NIMBY: Not In My Backyard). It therefore stands to reason that the amount of built-up surface already present in a grid cell has a sizeable impact on the likeliness of further built-up development, and that this impact is decidedly non-linear.

By design, the logit-based locational suitability functions, such as those used in this study, cannot extract information on how pre-existing built-up surface affects further built-up development.

To include the impact of pre-existing built-up surface in the overall modelling framework, Expected Top-ups (ET) factors have been included in the model. To compute these, transition matrices are created that sum occurrences of every initial built-up level and every outcome built-up level in the next decade. Subsequently, ETs are obtained by averaging the additional decennial built-up values per initial built-up level, expressed in percentage points increase. Some manual smoothing has been applied to reduce volatile changes in the distribution tails. These ETs are established separately for all continents, and continent-specific ETs are applied in the current implementation (Figure 10).

Figure 10. Currently implemented continent-specific Expected Top-ups (ET) values.



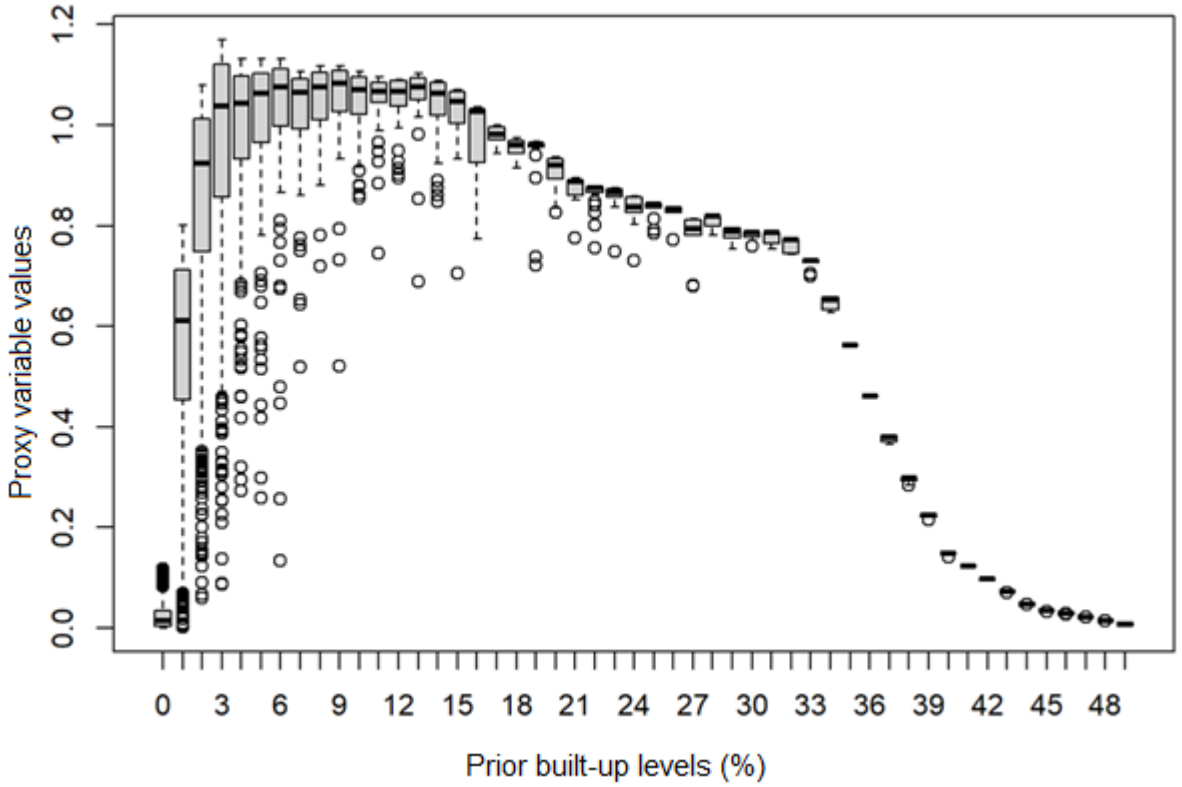
Source: CRISP model input, JRC analysis.

A note on the use of ETs is necessary here. In the reference application, a dynamic version of the ETs is being utilised. This entails a modified version of the empirically obtained ET values in which these values respond to the amount of space in an FA that is still available for built-up development. Effectively, this modification causes ETs at higher prior built-up levels to increase somewhat where space for built-up development is limited. This is explained further in Annex I.

4.4. Downscaling built-up development

The final proxy by which additional built-up values are downscaled is calculated as in equation (5), thus multiplying locational suitabilities, the built-up-level specific rebalancing factors, and ET values. The distributional results of this multiplication are given in Figure 11 for Nicaragua. This graph shows that: i) the combination of locational suitabilities, ii) their rebalancing and iii) the application of ETs create a rich pattern of outcomes, with particularly strong variation in low built-up levels, where ETs are high and the locational suitability varies considerably.

Figure 11. Distribution of the proxy variable used for downscaling built-up development values in Nicaragua.



Source: CRISP model output, JRC analysis.

The final probability for built-up development is subsequently used as a proxy to disperse the regional built-up claims. This is done in a straightforward scaling operation in which $\widehat{\Delta BU_i}$ acts as a proxy (7):

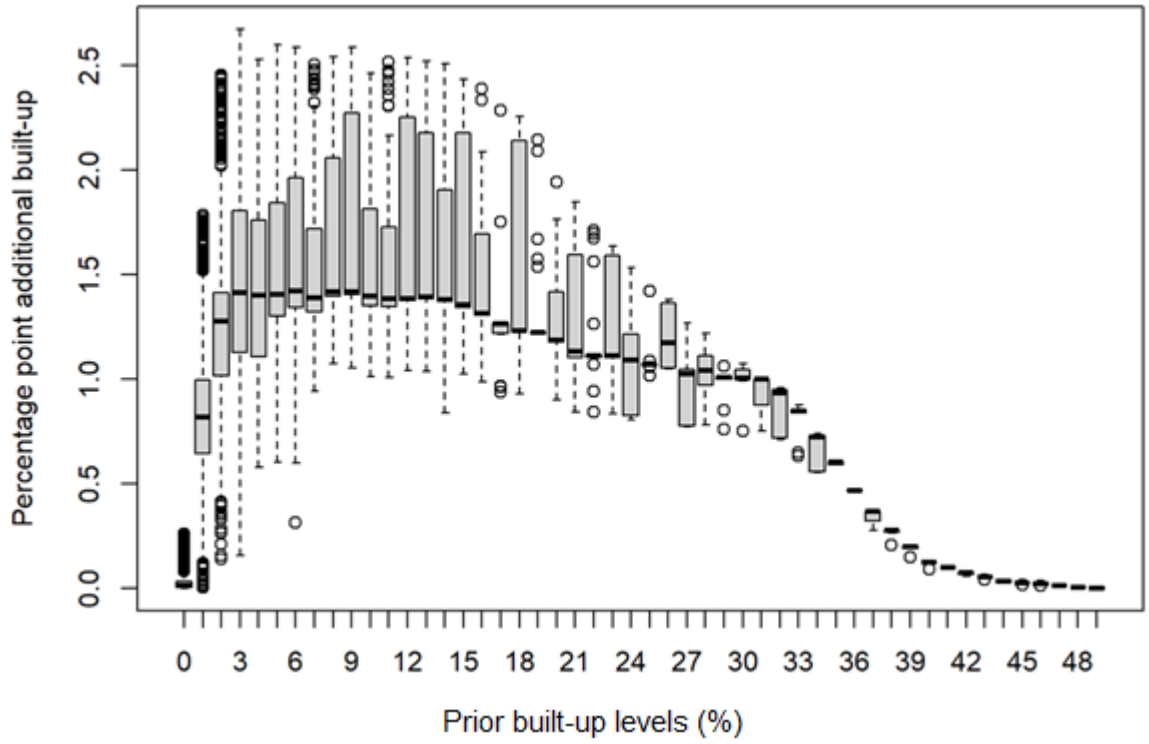
$$BU_{i,t} = BU_{i,(t-1)} + \Delta BU_{f,t} * (\widehat{\Delta BU_i} / \sum_{i=1}^{i \in F} \widehat{\Delta BU}_i), \quad (7)$$

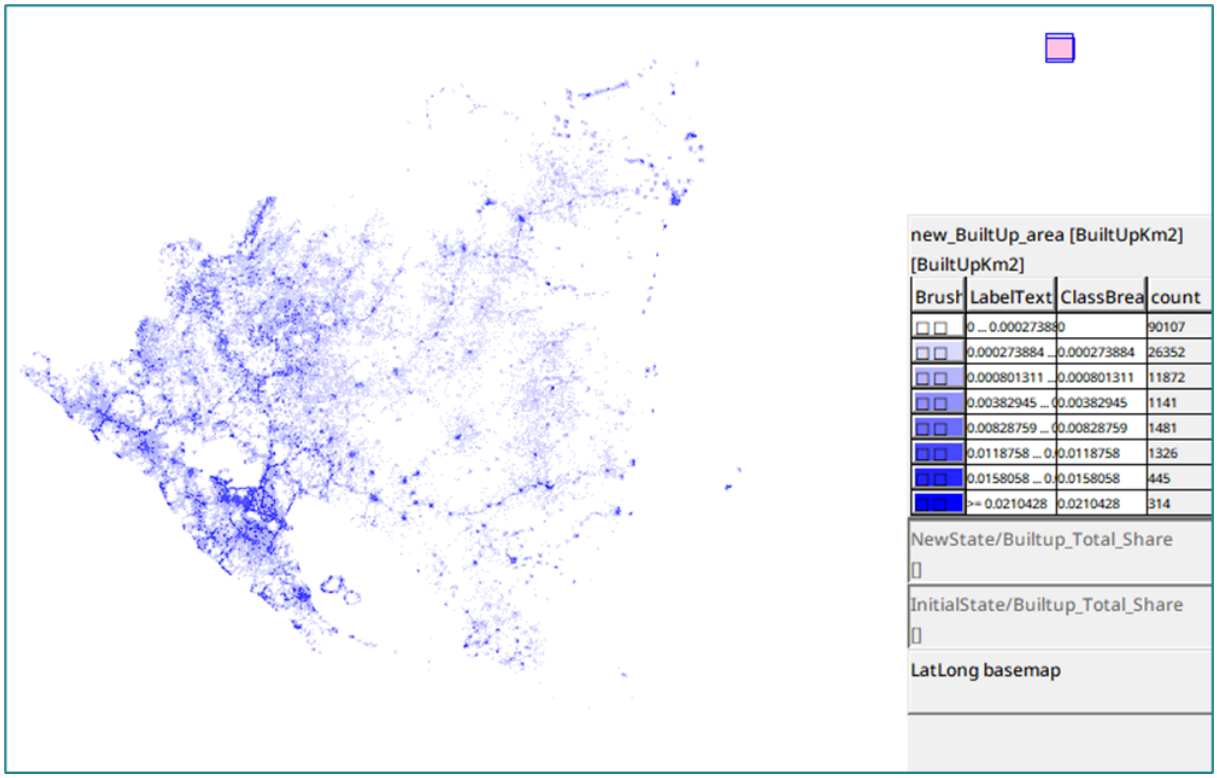
so that total built-up value *BU* in a cell *i* in model iteration *t* depends on

- prior built-up value $BU_{i,(t-1)}$ in the grid cell
- expected additional built-up development in the functional area ΔBU_f
- the proxy as summarised in Eq (1) and this section.

The result of this downscaling, without prior built-up, is shown graphically and geographically in Figure 12 for the case of Nicaragua.

Figure 12. Exemplary additional built-up values allocation results for Nicaragua by built-up levels (top) and on a map (bottom).





Source: CRISP model output, JRC analysis.

5. Population is added or relocated using universal rules, following built-up development and locational suitability

Population is distributed according to stylised, universal rules that depend on 1) the results of built-up development in the same model iteration; 2) locational suitability; and 3) a number of general time and location-specific limitations. We must concede that, ideally, the population distribution procedure is tied less stringently to built-up modelling than is currently the case (see e.g. Diogo *et al.*, 2023). To properly capture population changes, data is required that describes ground-truth, observed population changes with global coverage. Unfortunately, such population data is still elusive. The most reliable population data that is available, such as the GHS-POP layer provided by the GHSL team (Schiavina *et al.*, 2023), are the result of dasymetric mapping. This sort of population data relies on built-up fractions or built-up volume as a proxy. We make use of these GHS-POP data. In a later study, we plan to explore population changes in more detail in geographies where the needed ground-truth population grids are available.

The distribution of population change proceeds in four steps:

1. The initial population is reduced to accommodate population decline and facilitate net internal relocation. This reduction approximates migration out of grid cells. The reduced population is partially moved to an FA-specific population pool. Population is initially taken out of grid cells that are unattractive according to the logit transformation of locational suitabilities. We call this the *scaled reduction*.
2. If not enough people can be taken out of unattractive grid cells to obtain the expected number of outmigrants, a blanket percentage of additional people are drawn out of still populated grid cells. We call this the *proportional reduction*.
3. Then, built-up developments are populated based on the average prior built-up population density in neighbouring grid cells. A cell's neighbourhood is defined as all cells that have their centre in a 3 km radius from the centre of the origin-cell.
4. Any remaining population claim is met by iteratively assigning the population to grid cells according to locational suitability with an increasing scaling factor. This assignment is limited by the maximum allowed population density.

These steps are designed to allocate newcomers in an FA and slowly reallocate the population, taking into account population declines. In a baseline without exogenous population change or built-up development, this simulated relocation process causes slight population densification in the most suitable locations. The GHS-POP (Schiavina *et al.*, 2023) data serve as the basis for population modelling, with 2020 as the reference year. However, the aggregate populations in 2020 according to the GHSL layer do not match national populations in 2020 according to the population projections used in this study. To eliminate any inconsistencies between the aggregate population counts, a tailored set of population grids 1975 - 2020 were prepared that are fully consistent with the population totals according to the used WPP projections.

These reference gridded datasets provide population counts as continuous numbers, while the model assumes integer population counts. To ensure consistency between the reference and the modelled population estimates, the reference data are transformed to integer population counts while ensuring population totals are kept per FA. The transformation method is detailed in Section 5.5. In addition, we determine the number of persons per km^2 of residential built-up surface and truncate this to the maximum of 1 million people per km^2 of residential built-up surface (or one

person per m^2). Cases where the data exceed this limit are rare and implausible. This maximum threshold of people per built-up surface implies that the model imposes that people are only present in grid cells with some built-up surface. To ensure consistency, this maximum threshold is applied for both the reference data (1975 – 2020) and the modelled population data (2020 – 2100). In cascade, the reference gridded datasets are also adjusted to ensure that population never exceeds the imposed maximum, while maintaining population totals at the FA level.

5.1. Population maxima at the grid cell level

We determine the maximum number of persons per grid cell that the population may grow in a timestep. A timestep entails a decade, and is noted here as *t-1* to *t*. As a general rule, the model does not allow population densities to exceed 100,000 inhabitants per km^2. Furthermore, the model applies timestep-specific constraints. These constraints are derived from an extensive analysis of historical population changes according to the 2023 data release of the GHS-POP (Schiavina *et al.*, 2023). These constraints are:

- If a grid cell has the highest local population in a square of 3x3 in *t-1*, it can only increase by 3,500 people per km^2 built-up surface in *t*.
- If a grid cell is populated but it doesn't have the highest local population in a square of 3x3 in *t-1*, it can only increase in *t* to the local maximum in *t-1* + 2,500 people per km^2 built-up surface;
- If a cell is not populated in *t-1*, it can only increase by 350 people per km^2 built-up surface in *t*.

These constraints are multiplied by simultaneous built-up fractions; they are grid-cell specific, and recomputed in every timestep. For now we indicate these constraints as $c_{i,t}$.

5.2. Population pools at the Functional Area level

Every FA has a timestep-specific population pool ($O_{f,t}$) that always contains at least a minimum share of the FA's population. This pool is composed as (8):

$$O_{f,t} = \Delta P_{ft} + AR_{f,t} + I_{f,t} \quad (8)$$

containing:

- foreseen population change (ΔP_{ft}), as given by exogenous projections;
- a potential additional reduction ($AR_{f,t}$), which is set only in case $\Delta P_{ft} < 0$, so that $AR_{f,t} = \begin{cases} -1 * \Delta P_{ft} \; if \; \Delta P_{ft} < 0 \\ 0 \; if \; \Delta P_{ft} \geq 0 \end{cases}$; and
- a parametrised minimum percentage of the prior population that represents internal migration (I_f). In the reference application, this internal migration parameter is set at 1% so $I_{f,t} = 0.01 * P_{f,t-1}$.

If the regional population increases, the pool size equals the population growth + the portion of the regional population that is set to migrate within the region. If the regional population shrinks, a larger portion of the prior population is taken out of grid cells through *AR*, ensuring that the model

reallocates the desired internal migration rate. In the case of population decline in an FA, the population pool equals internal migrants.

The population pool is thus filled at least partially by population already present in the FA. This part of the population is drawn from individual grid cells in two steps. Let's represent the total population to be drawn into the population pool by $D_{f,t} = AR_{f,t} + I_{f,t}$.

Scaled reduction

To fill *D*, population is first taken out of locations that are scaled by unsuitability. This is done using the inverse of the logit probability function fitted in the calibration process and described in the built-up values allocation section. To this end, invoke a proxy for doing an unsuitability-scaled reduction of population *SRP* (9):

$$SRP_{i,t} = P_{i,t-1} * (1 - S_{i,t}), \tag{9}$$

which yields a higher number with more prior inhabitants and lower locational suitability[4]. Note that the logit transformation should ensure that $0 \leq S_{i,t} \leq 1$. Then downscale *D* into a grid-specific *D1* and a remaining population *P1* so that (10.1, 10.2):

$$D1_{i,t} = D_{f,t} * (SRP_{i,t} / \sum_{i=1}^{i \in F} SRP_{i,t}) \leq P_{i,t-1}, \tag{10.1}$$

$$P1_{i,t} = 0 \geq P_{i,t-1} - D1_{i,t}, \tag{10.2}$$

in which we ensure that the draw from grid cells is linear with the *SRP* proxy, but is capped, so that it cannot exceed the total population that was present in the grid cell. This is necessary to prevent intermediate negative population counts.

Additional proportional reduction

The requirement that a draw cannot exceed a grid cell's population implies that, in rare cases, the entire population necessary to fill *D* is not drawn in *D1*; i.e., $0 \geq D_{f,t} - \sum_{i=1}^{i \in F} D1_{i,t}$. We therefore do a second draw from the population *D2*. This second draw takes the form of a proportional reduction (11.1, 11.2):

$$D2_{i,t} = P1_{i,t} * (D_{f,t} - \sum_{i=1}^{i \in F} D1_{i,t} / \sum_{i=1}^{i \in F} P1_{i,t}), \tag{11.1}$$

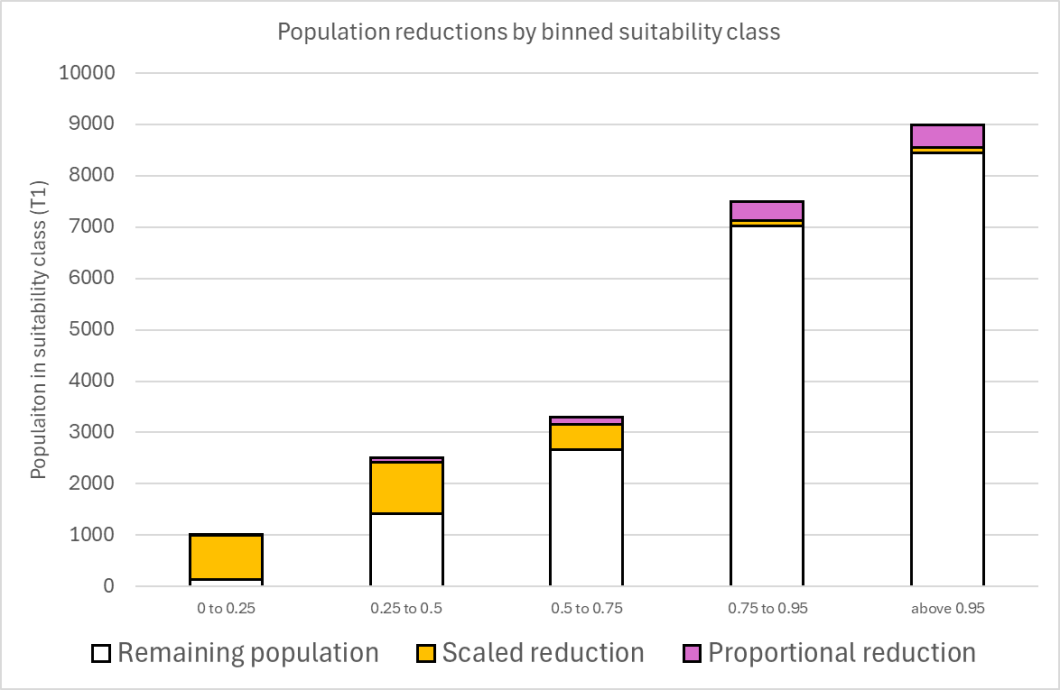
$$P2_{i,t} = P1_{i,t} - D2_{i,t}, \tag{11.2}$$

in which the potentially remaining regional reductions are taken proportionally out of the remaining population in *P1.*

[4] In earlier model versions, the amount of people drawn from unsuitable location also depends on the percentage of a grid cell that we maximally allow to be taken out of a grid cell. This percentage is governed through a parameter that is set to 0.9 (90%) in the reference application. It is advisable to set this parameter above 0.5, as that allows grid cells to become depopulated.

The below Figure 13 illustrates the potential distribution of reductions from the existing population (*t-1*). In this example, the majority of population reductions is drawn from cells with a low locational suitability at the step of the scaled reduction. However, another 5% of the population is drawn by proportional reduction. Practically, such a considerable proportional reduction only occurs in an FA undergoing a sizeable population decline.

Figure 13. Exemplary scaled and proportional population reductions by stylised locational suitability bins.



Source: schematic example, JRC analysis.

5.3. Allocating people to built-up developments

Next, we allocate population to built-up developments. Built-up developments in *t* are the change in built-up in a grid cell in the timestep at hand, so that built-up development $\Delta BU_{i,t} = BU_{i,t} - BU_{i,(t-1)}$ (see Eq. 7). These represent newly built-up fractions in every grid cell. Grid cells that were already partially built-upon in *t-1* can thus still have additional built-up surface assigned in *t*.

The amount of population that built-up developments receive depends on three factors:

- The number of people in the regional pool (*O*), as discussed in the previous section.
- The average prior built-up surface per capita, i.e. the prior population over the prior built-up surface in the 3x3 neighbourhood of the grid cell. This is defined mathematically as $wPD_{i,t} = \sum_{j=1}^{n} P_{i,t-1} W_{ij} / \sum_{j=1}^{n} BU_{i,t-1} W_{ij}$, with *W* being a spatial weighting matrix that is 1 if grid cells in *j* are within a 3x3 neighbourhood according to the Queen's case, and 0 otherwise.

The built-up surface per capita of these neighbourhoods is used to estimate additional population in built-up developments ($\hat{P}_{i,t}^{BU}$ in 12).

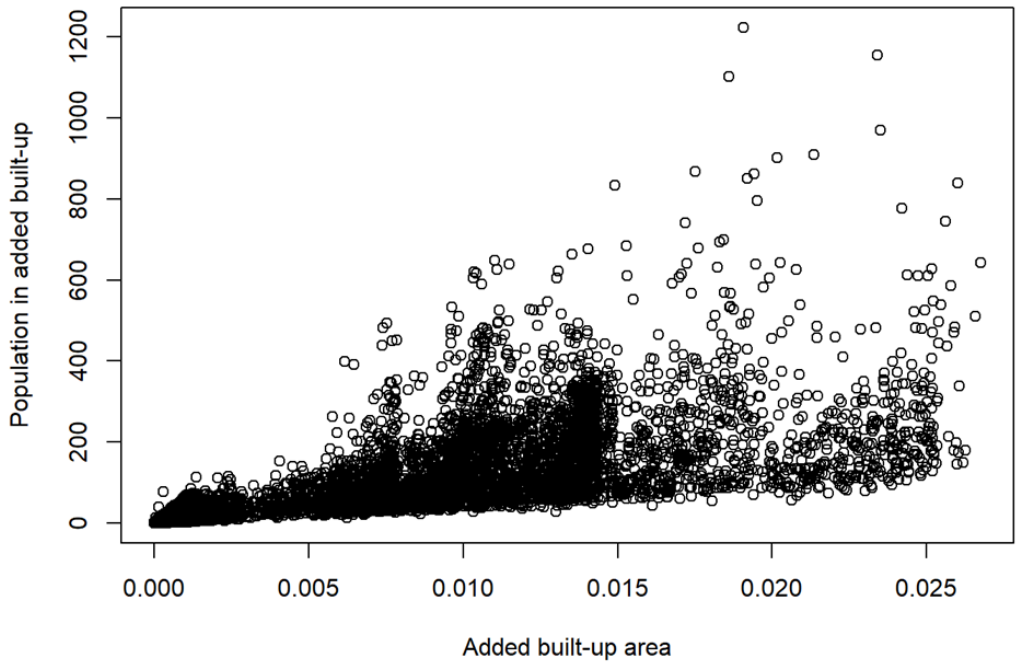
$$\hat{P}_{i,t}^{BU} = \Delta BU_{i,t} * wPD_{i,t}, \quad (12)$$

In some cases the summed estimated additional population in built-up developments overshoots[5] the population pool, ie $\sum_i^{i\in F} \hat{P}_{i,t}^{BU} > O_{f,t}$. In such cases $\hat{P}_{i,t}^{BU}$ is rescaled, leading to the definite additional population in built-up developments (13):

$$P_{i,t}^{BU} = \hat{P}_{i,t}^{BU} * ([O_{f,t}/\sum_i^{i\in F} \hat{P}_{i,t}^{BU}] \leq 1), \quad (13)$$

The results from equations (12) and (13) indicate that the remaining population plus the population just allocated to the grid cells with increased built-up value does not exceed the FA-specific projection, while at the same time retaining the geographical variation in population densities stemming from the local population density measure. In most cases, there remains some population to allocate to satisfy the FA-specific population projection. Figure 14 shows provisional results for Nicaragua by juxtaposing additional built-up value in a grid cell with the number of people allocated in a modelling timestep.

Figure 14. Provisional results from a modelling timestep in Nicaragua showing additional population allocated to the grid cells with increased built-up value as a function of the amount of built-up value in a grid cell.



Source: CRISP model output, JRC analysis.

[5] The implemented model contains an additional parameter ($0 \leq m \leq 1$) that optionally maximises the share of people in the regional pool that can be allocated to built-up developments. In the reference application this parameter is set to 1, so that there is no ad-hoc maximisation.

5.4. Allocating the remainder population while maintaining local constraints

Typically not all population in the FA-specific pool is allocated to built-up developments, so that after the operations from the previous section, a residual projected population remains for most FAs. More formally, let $P_{f,t}^{res}$ be the residual population $P_{f,t}^{res} = 0 \geq P_{f,t} - \sum_{i \in f}^{n}\left(P2_{f,t} + P_{i,t}^{BU}\right)$, in which $P_{f,t}$ is the FA and time-specific projected population, $P2_{i,t}$ is the population remaining after scaled and proportional reductions, as in Section 5.2, and $P_{i,t}^{BU}$ is the population allocated to built-up developments, as in Section 5.3.

An iterative process allocates this residual projected population proportionally to $L_{i,t}$, the untransformed locational suitability introduced in Section 4.1. Simultaneously, overall local population maximum constraints are imposed in this iterative process, ensuring that the modelled population does not exceed the previously established local and general maximum population levels. To achieve this, we first calculate the remaining vacancy for population per cell, which is the difference between the population from the previous step and the maximum allowed population per cell; i.e. the vacancy level $v_{i,t}$ in (14),

$$v_{i,t} = c_{i,t} - \left(P2_{i,t} + P_{i,t}^{BU}\right), \tag{14}$$

with $c_{i,t}$ indicating the absolute population constraints, specific for a grid cell and model iteration, resulting from the population maxima introduced in Section 5.1. Solving equation (14) yields the grid-specific vacancy level $v_{i,t}$, which we subsequently multiply with an indication of the locational suitability of the grid cell. This locational suitability derives from the values in $L_{i,t}$ (see Section 4.1) in the following manner. The values in *L* are transformed to ensure it only maintains positive numbers, with values that were above zero increased by 1, and the exponential taken from values that were below zero so that these values are effectively scaled between zero and 1. The resulting transformed values of *L* are subsequently multiplied by vacancy (15):

$$wv_{i,t} = v_{i,t} * \begin{cases} L_{i,t} + 1 \; if \; L_{i,t} \geq 0 \\ e^{L_{i,t}} if \; L_{i,t} < 0 \end{cases}. \tag{15}$$

The resulting weighted vacancy $wv_{i,t}$, a product of vacancy and locational suitability, is considered an abstract measure of attractivity. This attractivity factor is instrumental for the iterative allocation process here. We then calculate an initial scale factor *F0* by dividing the population that still needs to be allocated by the summed attractivity factors in the FA, i.e. (16):

$$F0_{f,t} = P_{f,t}^{res} / \sum_{i \in f}^{n} wv_{i,t}, \tag{16}$$

which expresses the number of people to allocate $P_{f,t}^{res}$, per unit of attractivity $wv_{i,t}$.

Subsequently iterations are needed to optimise the distribution of population given attractivity. This iterative process has crude similarities with an iterative proportional fitting procedure. Per iteration, an adapted scale factor is fed into the loop. Starting from the initial scale factor *F0*, then *F1, F2, …. Fn*,

1. proportional population is distributed by multiplying the grid-cell proportionality with the input regional scale factor, $\Delta\hat{P}n_{i,t} = wv_{i,t} * Fn_{f,t} \;\; (i \in f)$

2. proportional population is capped so that it does not exceed the vacancy, $\Delta Pn_{i,t} = \Delta \hat{P}n_{i,t} \leq c_{i,t}$
3. capped proportional population is added to the population never moved and allocated to new built-up areas, yielding realised total population, $Pn_{i,t} = P2_{i,t} + P_{i,t}^{BU} + \Delta Pn_{i,t}$
4. a regional rescaling factor is estimated as the factor difference between expected, exogenous, regional population and the population realised in this iteration, so that $\hat{F}n_{f,t} = P_{f,t} / \sum_{i \in f}^{n} Pn_{i,t}$; and finally,
5. the regional scale factor is multiplied with the rescaling factor, and subsequently passed to the next iteration so that $Fn_{f,t} = F0_{f,t} * \hat{F}n_{f,t}$. The process starts again from 1) with this new factor.

Effectively, this iterative procedure starts by distributing the remaining population over the product of population vacancy and locational suitability. In some places, this will cause an overshoot vis-a-vis vacancies. Capping on vacancy means that, usually, the input scale factor *F0* does not lead to allocation of the entire remainder population. Thus, an iterative rescaling is needed in which the proportional populations keeps producing larger overshoots, until the sum of the vacancy-capped population (i.e., $\sum_{i \in f}^{n} Pn_{i,t}$) equals the wanted remainder population size $P_{f,t}$.

5.5. Transform to integer population count

The iterative increase and then decrease of population at the grid cell level yields an optimised population distribution, equal to the aggregate projected population and respecting the imposed constraints. However, the population counts are expressed as a floating number, so that fractions of a person may be present in a grid cell. We subsequently transform these fractions into an integer number of people using a simple workflow:

1. transform the floating population to an integer
2. calculate n as the total population lost after transformation, summing total continuous population in the FA minus the summed integer population in the FA.
3. rank grid cells by the size of the fraction lost because of transformation to integer
4. select the n grid cells with the highest rank from 3) and add a person to each.

This process ensures integer population while maintaining the projected population totals.

6. Conclusions

The presented model is a functional method by which national projections of population change are downscaled to the 1 km grid level. This is done by first disaggregating national projections to regional units. In the presented application for the 2025 World Urbanisation Prospects, national projections are downscaled to so-called Functional Areas (FAs). Those FAs were defined for this study specifically with the aim of providing cohesive subnational units of approximately comparable sizes. We assume the defined FAs represent relatively homogeneous populations; testing is necessary to verify if indeed FA groupings capture demographically more homogenous populations than other subnational units. After downscaling population to the FA level, built-up surface projections at the FA-level are computed using expected changes in built-up surface per capita as an intermediary.

Subsequently, regional built-up surface projections up to year 2100 can be downscaled to the gridded, 1 km resolution built-up fractions in a statistically informed approach. These fractions are subsequently used to guide the allocation of population change. The built-up model is informed through parameters estimated in logit models. The population model is based on assumptions and observed maximum changes. Statistical models of global local population change are still evasive, as the available requisite population grids are a product of dasymetric mapping methods and cannot be considered ground-truth. Despite these limitations, the model is able to produce plausible patterns and trends of population change.

The built-up model relies on logit models that explain the binary presence of built-up grid cells. This modelling approach yields robust results, and the custom AutoGLM package that was used for modelling is a tested and reliable method to evaluate locational suitability. One limitation of this logit approach is that the prior amount of built-up surface cannot be included as a factor in the locational suitability model. This limitation has been circumvented in CRISP by introducing Expected Top-up (ET) functions that describe the average additional built-up fraction per prior built-up level. Recent tests with so-called beta regressions indicate that such regressions may be reliably used to estimate locational suitability on built-up fractions or even the absolute change in built-up fraction; this will be explored for future improvements in the modelling.

As of now the grid-level model is being validated. In this approach, the model is used to reproduce observed built-up expansion and population change between 1975 and 2020 according to GHSL products. The model's accuracy in built-up and population allocation is then compared with other approaches to downscale population change. As a first step in this validation procedure, results of built-up development simulated through the current chain of assumptions are being compared with results from other statistical approaches including the beforementioned beta regressions as well as cubist submodels. For specific geographies (the EU, Japan and Korea), census grids are available for multiple points in time. These census grids are used to verify how well the model stacks up to ground truth, rather than dasymetrically mapped population changes. The results of validation will be published in a separate follow-up paper. Results from validation will guide improvements to the CRISP model, and are critical in a planned version of the CRISP model dedicated entirely to Europe.

References

Andree, B.P.J. and Koomen, E. (2017) *Calibration of the 2UP model*. Vrije Universiteit Amsterdam (Spinlab Research Memorandum).

CIESIN (2018) *Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11 (Version 4.11)* . Available at: https://doi.org/https://doi.org/10.7927/H4JW8BX5.

Dijkstra, L. and Jacobs-Crisioni, C. (2023) "Developing a definition of Functional Rural Areas in the EU." Ispra (Italy): European Commission. Available at: https://publications.jrc.ec.europa.eu/repository/handle/JRC135599.

Diogo, V. *et al.* (2023) "Integrated Spatial Simulation of Population and Urban Land Use: a Pan-European Model Validation," *Applied Spatial Analysis and Policy*, 16(4), pp. 1463–1492. Available at: https://doi.org/10.1007/s12061-023-09518-x.

Dottori, F. *et al.* (2016) *Flood hazard map of the World - 100-year return period. , European Commission, Joint Research Centre (JRC) [Dataset]*. Available at: http://data.europa.eu/89h/jrc-floods-floodmapgl_rp100y-tif (Accessed: August 9, 2025).

ESA (2024) *Copernicus Global Digital Elevation Model*. Available at: https://doi.org/https://doi.org/10.5069/G9028PQB.

EU-CLMSI (2015) *Land Cover 2015-2019 (raster 100 m), global, annual - version 3*. Available at: https://doi.org/https://doi.org/10.2909/c6377c6e-76cc-4d03-8330-628a03693042.

Ferdinand, P., Andree, B.P.J. and Koomen, E. (2021) *Revised calibration of the 2UP model; analysing change and regional variation*. Amsterdam. Available at: https://hdl.handle.net/1871.1/c0bbdc8b-ef56-402b-8a4c-4847818cdb6e (Accessed: September 22, 2025).

Freire, S. *et al.* (2016) "Development of new open and free multi-temporal global population grids at 250 m resolution," in *Proceedings of the 19th AGILE conference on Geographic Information Science. Helsinki, Finland, June 14-17, 2016*.

Jacobs-Crisioni, C., Kompil, M. and Dijkstra, L. (2023) "Big in the neighbourhood: Identifying local and regional centres through their network position," *Papers in Regional Science*, 102(2), pp. 421–457. Available at: https://doi.org/https://doi.org/10.1111/pirs.12727.

ObjectVision (2023) *Geo data and model server (GeoDMS)*. Available at: https://github.com/ObjectVision/GeoDMS/wiki (Accessed: August 28, 2023).

OSM (2023) *Continental OpenStreetMap network data*. Available at: https://download.geofabrik.de/ (Accessed: January 1, 2023).

Pesaresi, M. *et al.* (2024) "Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data," *International Journal of Digital Earth*, 17(1), p. 2390454. Available at: https://doi.org/10.1080/17538947.2024.2390454.

Pesaresi, M. and Politis, P. (2023) "GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030)." European Commission, Joint Research Centre (JRC).

Schiavina, M. *et al.* (2023) *GHS-POP R2023A - GHS population grid multitemporal (1975-2030).* Available at: https://doi.org/https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE.

UN (2024) *World Population Prospects 2024, Data Sources*. New York.

van der Wielen, T. and Koomen, E. (2024) *Calibration of the CRISP model: A global assessment of local built-up area presence*, *A global assessment of local built-up area presence*. VU University/ SPINlab.

List of abbreviations and definitions

Abbreviations	Definitions
CRISP	Cities and Rural Integrated Spatial Projections
WUP	World Urbanisation Prospects
SSP	Shared Socio-economic Pathways
GIS	Geographic Information System
WPP	World Population Prospects (UN national population projections)
FA(s)	Functional Area(s)
GHSL	Global Human Settlement Layer
GHS-BUILT	GHS built-up grids (v2023A)
GHS-POP	GHS population grids (v2023A)
ET	Expected Top-ups
OSM	Open Street Map public road network data
2UP	Towards an Urban Preview model (preceding CRISP)
NIMBY	'Not In My Backyard' stance towards development

List of figures

List of tables

Annexes

Annex 1. Dynamically weighted Expected Top-ups

The implemented Expected Top-up (ET) values reflect the distribution of built-up development given historical, presumably mostly unconstrained increases in aggregate built-up land. The projections applied in the reference application indicate very rapid population growth and even more rapid urbanisation in parts of the world. In some cases this rapid growth consumes a large part of the open space in a functional area. As mentioned in the report, the ET mechanism has been adapted somewhat to allow for more flexibility if built-up land is expected to occupy a large percentage of available land.

The implemented static ETs would cause entire regions to become developed mostly with mid-range built-up fractions. We expect that, in the case of considerable pressure of built-up land on total available area, open space is treated more frugally, redirecting urban development to concentrated locations. As a project-specific assumption, the ETs have therefore been modified with a dynamic component. This dynamic specification hinges on the definition of available space in a function area $A_{f,t+1}$ (17):

$$A_{f,t+1} = BU_{f,t+1} \,/S_f^{compacted}\,, \tag{17}$$

which depends on the total expected built-up area (*BU*) and the total amount of space available in the compacted domain in a functional area, $S_f^{compacted}$. We utilise *A* to compute a dynamic weight *AW* (18):

$$AW_{i,t+1} = \left(\%BU_{i,t} * 100\right)/A_{i\,\epsilon\,f,t+1}, \tag{18}$$

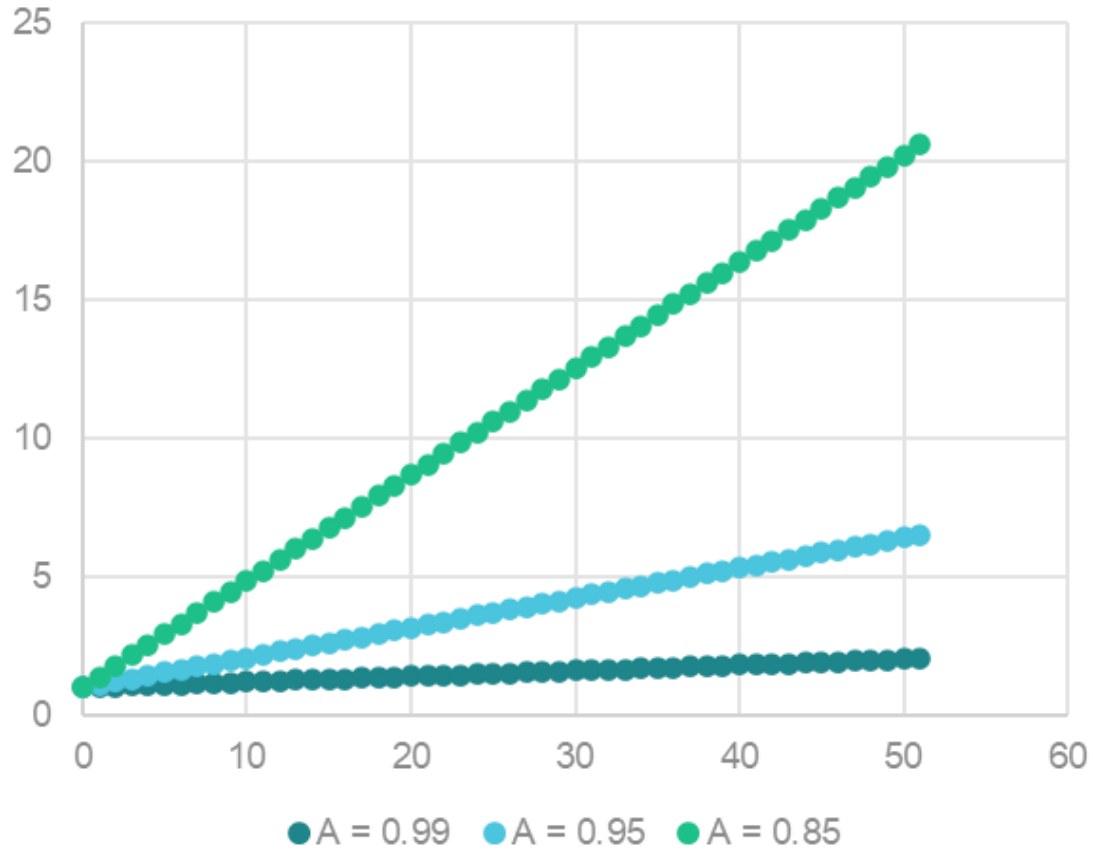
We derive the original Expected Top-up given continental ET values and the gridcell-specific built-up fraction as $ET_{i,t+1} = f(BU_{i,t})$. We then finally modify this expected value with the available space weight so that (19):

$$DET_{i,t+1} = \left(1 + AW_{i,t+1}\right) * ET_{i,t+1}, \tag{19}$$

Subsequently, $ET_{i,t+1}$ is replaced by $DET_{i,t+1}$ in the proxy $\widehat{\Delta BU_i}$ discussed in Section 4.4.

The dynamic weighting of ET values effectively increases ET values for grid cells with higher prior built-up levels for all cases where the available space is below 1. Figure 15 shows the distribution of available space weights for different values of Available space (A).

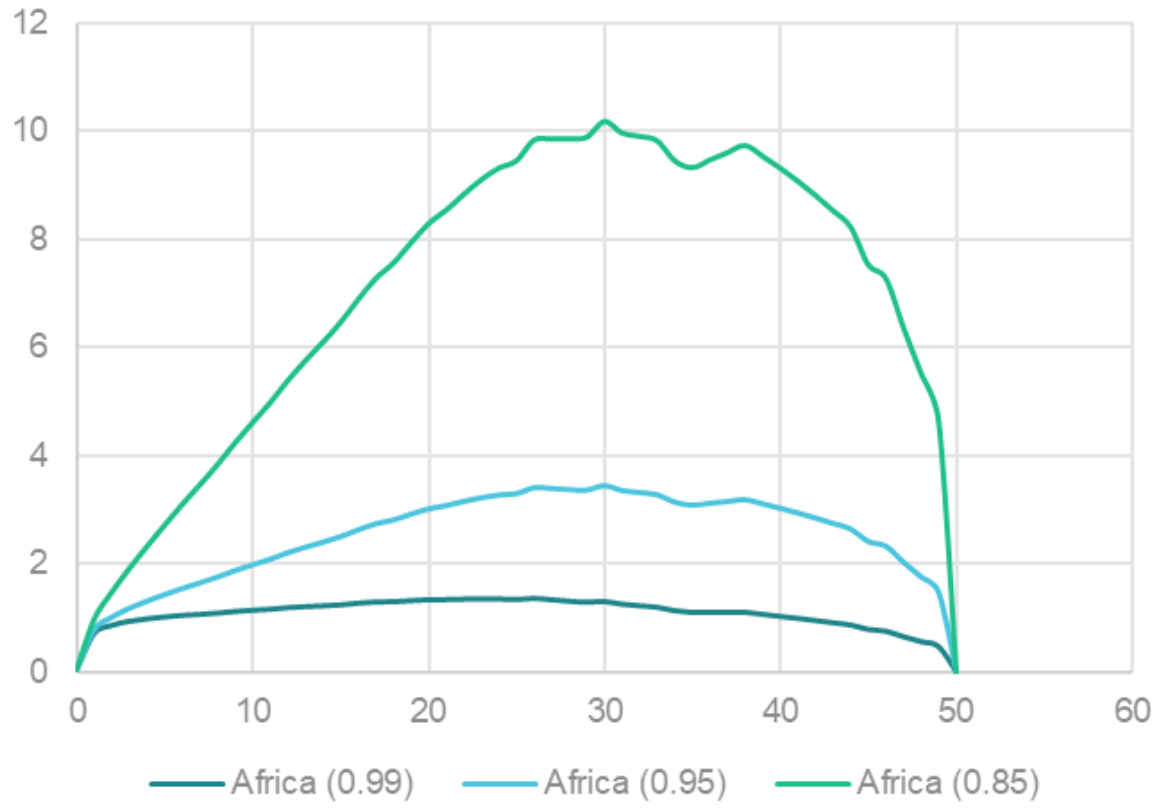
Figure 15. Available space weights (AW, y-axis) for different levels of prior built-up development (BU, x-axis) and available space (A, colours).



Source: CRISP model input, JRC analysis.

Multiplying ET values with these available space weights causes a notable general value increase, see Figure 16. A general value increase does not affect the downscaling, as the weighted ETs are only used as a proxy variable in a scaled distribution.

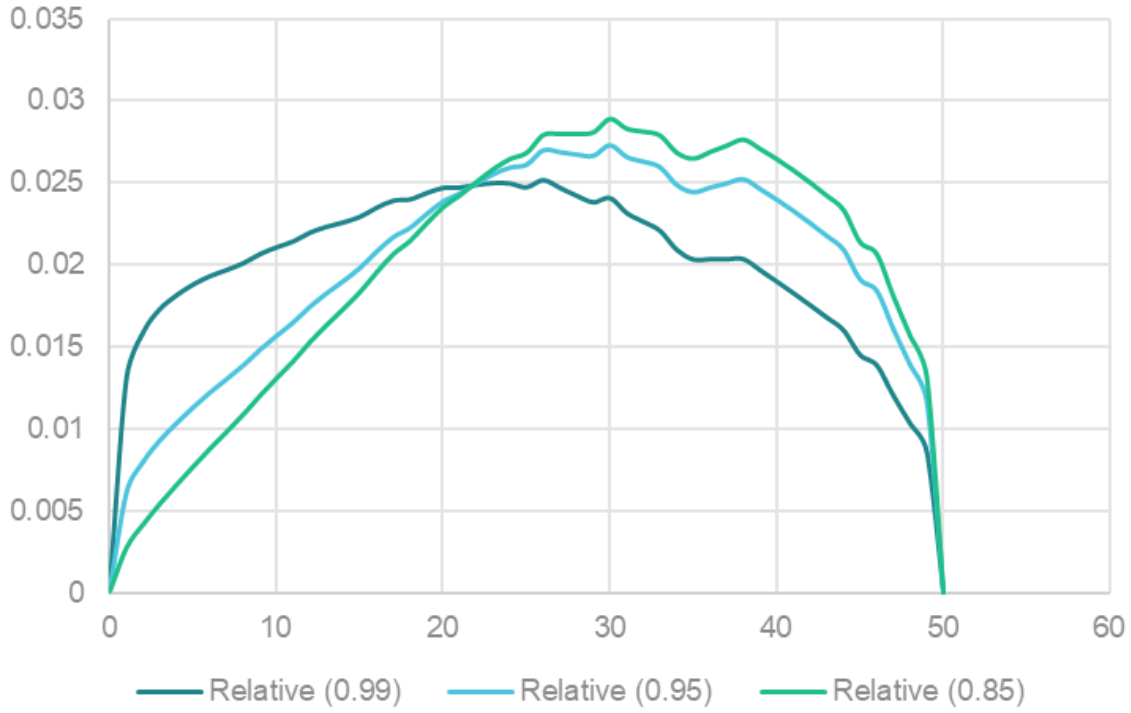
Figure 16. Dynamic ET values (DET, y-axis) multiplied with tentative ET values for different prior built-up levels (BU x 100, x-axis) for Africa, given available space (A, colours).



Source: CRISP model input, JRC analysis.

However, the weighting also causes a rearrangement of the distribution. This is more obvious when scaling back the above three weighted distributions, so that for all three lines the sum of all values is 1 (see Figure 17). There clearly the weighting causes increasing emphasis on the right-hand side of the current built-up range. In turn, this nudges the model to concentrate more of the additional built-up development in areas that already have considerable levels of development.

Figure 17. Dynamic ET values (DET, y-axis) for Africa by prior built-up values (BU x 100, x-axis) given available space (A, colours), rescaled so that the sum of all values in a line is 1. These lines show an increasing emphasis on high prior built-up values in case of limited available space.



Source: CRISP model input, JRC analysis.

Annex 2. CRISP calibration input data

Multitemporal built-up surface data

The GHSL multitemporal built-up surface spatial raster dataset (GHS-BUILT-S; Pesaresi and Politis, 2023) depicts the distribution of built-up surfaces, expressed as number of square metres. Data are spatial-temporal interpolated or extrapolated from 1975 to 2030 in 5 years intervals (European Commission. Joint Research Centre., 2023; Pesaresi et al., 2024).

The GHS-BUILT-S spatial raster dataset (Table 4) represents built-up surface area per grid cell. The GHS-BUILT-S_NRES spatial raster dataset (Table 5) represents non-residential built-up surface area per grid cell. The difference between the total and non-residential built-up area produces the residential built-up area per grid cell.

Table 4. Main characteristics of GHS- BUILT-S R2023A spatial raster dataset

Dataset	GHS-BUILT-S_GLOBE_R2023A spatial raster
Description	Square metres of total built-up surface by epoch
Coordinate system	World Mollweide (EPSG:54009), WGS84 (EPSG:4326)
Spatial resolution	10m, 100m, 1 km, 3 arcsec, 30 arcsec
Temporal resolution	12 epochs, 1975 – 2030, 5 years timestep
Format	Tiff
Extent	Globe
Dataset size	Up to ~3GB per raster
Legend	The amount of square metres of built-up surface in the cell. Range and NoData value dependent on dataset resolution.
Source	https://ghsl.jrc.ec.europa.eu/download.php?ds=bu

Table 5. Main characteristics of GHS- BUILT-S_NRES R2023A spatial raster dataset

Dataset	GHS-BUILT-S_NRES_GLOBE_R2023A spatial raster
Description	Square metres of non-residential built-up surface by epoch
Coordinate system	World Mollweide (EPSG:54009), WGS84 (EPSG:4326)
Spatial resolution	10m, 100m, 1 km, 3 arcsec, 30 arcsec
Temporal resolution	12 epochs, 1975 – 2030, 5 years timestep
Format	Tiff
Extent	Globe
Dataset size	Up to ~1GB per raster
Legend	The amount of square metres of non-residential built-up surface in the cell. Range and NoData value dependent on dataset resolution.
Source	https://ghsl.jrc.ec.europa.eu/download.php?ds=bu

Multitemporal population data

The GHSL multitemporal population spatial raster dataset (GHS-POP; Schiavina, Freire, and MacManus, 2023) depicts the distribution of residential population, expressed as the number of people per cell (Table 6).

Residential population estimates between 1975 and 2020 in 5-year intervals and projections to 2025 and 2030 derived from CIESIN GPWv4.11 (Center For International Earth Science Information Network-CIESIN-Columbia University, 2018) were disaggregated from census or administrative units

to grid cells, informed by the distribution, volume, and classification of built-up as mapped in the GHSL data per corresponding epoch (Freire et al., 2016).

Table 6. Main characteristics of GHS-POP R2023A spatial raster dataset

Dataset	GHS-POP_GLOBE_R2023A spatial raster
Description	Population counts by epoch
Coordinate system	World Mollweide (EPSG:54009), WGS84 (EPSG:4326)
Spatial resolution	100m, 1 km, 3 arcsec, 30 arcsec
Temporal resolution	12 epochs, 1975 – 2030, 5 years timestep
Format	Tiff
Extent	Globe
Dataset size	Up to ~11GB per raster
Legend	The absolute number of inhabitants of the cell. NoData [-200]
Source	https://ghsl.jrc.ec.europa.eu/download.php?ds=pop

Inland water

The inland water spatial dataset (INLAND) indicates the (Boolean) presence of water. The INLAND dataset is produced in three steps. First, polygons of land areas obtained from Open Street Map (for 2021) are used as a global land mask. The polygons are rasterized to 1 m resolution in World Mollweide projection (ESRI:54009) and aggregated to 10 m resolution. Secondly, the binary water surface layer (WATER) is derived from the GHSL land fraction product in 100 m resolution (GHS-LAND; Pesaresi and Politis, 2022) by selecting grid cells with land fraction lower than 50% (5 000m^2 for 100 m resolution product): *WATER=GHS-LAND < 5000*. The WATER layer is clipped to the global land extent using the OSM-derived mask of land fraction, resulting in the binary INLAND layer.

The INLAND layer is vectorised to a spatial polygon dataset and reprojected to WGS84 projection (Table 7).

Table 7. Main characteristics of INLAND spatial dataset

Dataset	INLAND
Description	Inland water spatial polygon
Coordinate system	WGS84 (EPSG:4326)
Spatial resolution	N/A
Temporal resolution	2018
Format	GeoPackage
Extent	Globe
Dataset size	4 GB
Legend	Presence (1) of inland water.
Source	Upon request

Coastline

The coastline spatial polygon dataset (COASTLINE) represents the coastline of ocean waters. It is derived from the GHSL land fraction product in 100 m resolution (GHS-LAND; Pesaresi and Politis, 2022), greater or equal to 50% (5 000m^2 for 100 m resolution product): *LAND = GHS-LAND >= 5000*.

The COASTLINE layer is vectorised to a spatial polygon dataset and reprojected to WGS84 projection (Table 8).

Table 8. Main characteristics of COASTLINE spatial dataset

Dataset	COASTLINE
Description	Coastline of ocean waters spatial polygon
Coordinate system	WGS84 (EPSG:4326)
Spatial resolution	-
Temporal resolution	2018
Format	GeoPackage
Extent	Globe
Dataset size	250 MB
Legend	Presence (1) of coastline
Source	Upon request

Elevation and slope

Copernicus GLO-90 Digital Elevation Model (DEM) (European Space Agency and Airbus, 2022) represents the global surface of the Earth (including infrastructure and vegetation) at the resolution of 3 arc seconds (Table 9). The GLO-90 dataset is a composite of WorldDEM, a product of the radar satellite data acquired during the TanDEM-X Mission, and other DEMs.

Table 9. Main characteristics of GLO-90 spatial raster dataset

Dataset	GLO-90
Description	Digital Surface Model of Earth
Coordinate system	WGS84 (EPSG:4326)
Spatial resolution	3 arcsec
Temporal resolution	2010-2015
Format	Tiff
Extent	Globe
Dataset size	~100 GB
Legend	Elevation in meters
Source	https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model https://copernicus-dem-30m.s3.amazonaws.com/readme.html

Flood prone area

Flood hazard map of the World (Dottori *et al.*, 2016) depicts flood prone areas at global scale for flood events with 100-year return period (Table 10). Resolution of the product is 30 arcseconds. Cell values indicate water depth (in m). The map can be used to assess flood exposure and risk of population and assets.

Table 10. Main characteristics of GloFAS – flood hazard map of the World with a 100-year return period

Dataset	GloFAS – flood hazard
Description	Flood hazard map of the World - 100-year return period
Coordinate system	WGS84 (EPSG:4326)
Spatial resolution	30 arcsec
Date issued	2016

Format	Tiff
Extent	World
Dataset size	65 MB
Legend	Water depth [m]
Source	https://data.jrc.ec.europa.eu/dataset/jrc-floods-floodmapgl_rp100y-tif

Earthquake intensity

The Global Earthquake Model (GEM) Global Seismic Hazard Map v2023.1 (Johnson et al., 2023) depicts the geographic distribution of the Peak Ground Acceleration (PGA) with a 10% probability of being exceeded in 50 years, computed for reference rock conditions (shear wave velocity, Vs30 , of 760-800 m/s). The map was created by collating maps computed using national and regional probabilistic seismic hazard models developed by various institutions and projects, in collaboration with GEM Foundation scientists.

Table 11. Main characteristics of GEM v2023.1 raster dataset

Dataset	GEM v2023.1
Description	GEM Global Seismic Hazard Map
Coordinate system	WGS84 (EPSG:4326)
Spatial resolution	3 arcmin
Date issued	2023
Format	Tiff
Extent	World
Dataset size	165 MB
Legend	PGA Range
Source	https://zenodo.org/records/8409647

Landslide prone areas

The GFDRR Global Landslide hazard map (The World Bank, 2020) comprises gridded maps of estimated annual frequency of significant landslides (occurred in a populated place; at least greater than 100 m2), triggered by seismicity and rainfall.

The Global Landslide Hazard Maps present the mean or median annual landslide hazard assessment for the period 1980 – 2018. Raster values represent the modelled average annual frequency of significant landslides per sq. km (Table 12).

Table 12. Main characteristics of the GFDRR Global Landslide hazard map

Dataset	The Global Landslide hazard map
Description	GEM Global Seismic Hazard Map
Coordinate system	WGS84 (EPSG:4326)
Spatial resolution	30 arcsec
Date issued	2023
Format	Tiff
Extent	land between 60°S and 72°N
Dataset size	~60 MB per map
Legend	average annual frequency of landslides
Source	https://datacatalog.worldbank.org/search/dataset/0037584/Global-landslide-hazard-map

Full list of data used

Center For International Earth Science Information Network-CIESIN-Columbia University, 'Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11', Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2018.

Dottori, F., P. Salamon, A. Bianchi, L. Alfieri, F.A. Hirpa, and L. Feyen, 'Development and Evaluation of a Framework for Global Flood Hazard Mapping', *Advances in Water Resources*, Vol. 94, August 2016, pp. 87–102.

European Commission. Joint Research Centre., GHSL Data Package 2023., Publications Office, LU, 2023.

European Commission Joint Research Centre, 'INFORM Risk Index 2024', 2024.

European Space Agency and Airbus, 'Copernicus DEM', European Space Agency, 2022.

Freire, S., K. MacManus, M. Pesaresi, E. Doxsey-Whitfield, and J. Mills, 'Development of New Open and Free Multi-Temporal Global Population Grids at 250 m Resolution', *Proc. of the 19th AGILE Conference on Geographic Information Science*, Vol. 250, Vol. 250, AGILE, Helsinki, Finland 14–17 June, 2016.

Johnson, K., M. Villani, K. Bayliss, C. Brooks, S. Chandrasekhar, T. Chartier, Y.-S. Chen, et al., 'Global Seismic Hazard Map', Zenodo, Octobe023.

Pesaresi, M., and P. Politis, 'GHS-BUILT-S R2023A - GHS Built-up Surface Grid, Derived from Sentinel2 Composite and Landsat, Multitemporal (1975-2030)', European Commission, Joint Research Centre (JRC), April 25, 2023.

———, 'GHS-LAND R2022A - Land Fraction as Derived from Sentinel2 Image Composite (2018) and OSM Data', European Commission, Joint Research Centre (JRC), June 25, 2022.

Pesaresi, M., M. Schiavina, P. Politis, S. Freire, K. Goch, and J.H. Uhl, 'Advances on the Global Human Settlement Layer by Joint Assessment of Earth Observation and Population Survey Data', *International Journal Of Digital Earth*, 2024.

Schiavina, M., S. Freire, and K. MacManus, 'GHS-POP R2023A - GHS Population Grid Multitemporal (1975-2030)', European Commission, Joint Research Centre (JRC), April 25, 2023.

Schiavina, M., M. Melchiorri, and M. Pesaresi, 'GHS-SMOD R2023A - GHS Settlement Layers, Application of the Degree of Urbanisation Methodology (Stage I) to GHS-POP R2023A and GHS-BUILT-S R2023A, Multitemporal (1975-2030)', European Commission, Joint Research Centre (JRC), April 25, 2023.

The World Bank, *The Global Landslide Hazard Map*. Final Project Report, 2020.

Trigg, M.A., C.E. Birch, J.C. Neal, P.D. Bates, A. Smith, C.C. Sampson, D. Yamazaki, et al., 'The Credibility Challenge for Global Fluvial Flood Risk Analysis', *Environmental Research Letters*, Vol. 11, No. 9, September 1, 2016, p. 094014.

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
— at the following standard number: +32 22999696,
— via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

Scan the QR code to visit:

The Joint Research Centre: EU Science Hub
https://joint-research-centre.ec.europa.eu

Publications Office
of the European Union