

HANZE: Historical Analysis of Natural Hazards in Europe

Database documentation

Dominik Paprotny

Research Department Transformation Pathways, Potsdam Institute for Climate Impact Research,
Telegrafenberg A 31, 14412 Potsdam, Germany.

Correspondence to: Dominik Paprotny (dominik.paprotny@pik-potsdam.de)

Document version: 2.1

Summary. HANZE database, or ‘Historical Analysis of Natural Hazards in Europe’, aims to provide information on exposure to natural hazards for 42 European countries and territories from 1870 to 2020 in 100 m resolution. The database was constructed using high-resolution maps of present land use and population, a large compilation of historical statistics, and relatively simple and explicit models and disaggregation techniques. It is accompanied by a compilation of past damaging floods in Europe, which contains information on dates, locations and losses for 2519 events (1870–2020). Demographic and economic data encompassed in HANZE allow to ‘normalize’ information on losses due to natural hazards by considering price inflation as well as changes in population, production and wealth. It can be utilized to study changes in exposure, vulnerability and risk to various natural hazards. See also the project website: <https://naturalhazards.eu/>

Table of Contents

Citation.....	4
Acknowledgements	4
Disclaimer.....	5
1. Introduction.....	6
2. Basic characteristics of the database	6
2.1. Data format.....	6
2.2. Coverage	6
2.2.1. Spatial coverage.....	6
2.2.2. Temporal coverage	8
2.2.3. Thematic coverage.....	8
2.3. Connection between HANZE-Exposure and HANZE-Events.....	9
2.4. List of files	9
2.5. Code.....	12
3. HANZE-Exposure: methodology and detailed contents	14
3.1. Outline	14
3.2. Baseline maps.....	15
3.2.1. Administrative boundaries.....	15
3.2.2. Land cover/use	15
3.2.3. Soil sealing.....	17
3.2.4. Population	18
3.3. Input database of historical statistics	23
3.3.1. Structure of input dataset files.....	23
3.3.2. NUTS 3 regions	25
3.3.3. Total population	27
3.3.4. Urban population.....	27
3.3.5. Mean number of persons per household	28
3.3.6. Land use structure	28
3.3.7. GDP and its composition	29
3.3.8. Wealth and its composition	29
3.3.9. Conversion of original damage values in monetary terms	30
3.4. Land use and population distribution modelling	32
3.4.1. Sub-regional population redistribution.....	32
3.4.2. Urban population and urban areas (CLC 111-112)	35
3.4.3. Rural population redistribution	37
3.4.4. Industrial or commercial units (CLC 121)	38

3.4.5.	Road and rail networks and associated land (CLC 122)	38
3.4.6.	Airports (CLC 124)	38
3.4.7.	Construction sites (CLC 133).....	39
3.4.8.	Green urban areas, sport and leisure facilities (CLC 141 and 142).....	39
3.4.9.	Port areas, mineral extraction sites, dump sites (CLC 123, 131 and 132).....	40
3.4.10.	Agriculture (CLC 211-244)	40
3.4.11.	Burnt areas (CLC 334).....	42
3.4.12.	Vegetated natural areas (CLC 311–324, 333 and 411–422).....	42
3.4.13.	Areas covered by water, incl. intertidal flats (CLC 423 and 511–523)	43
3.4.14.	Other natural areas (CLC 331–332 and 335)	43
3.4.15.	Special cases	43
3.4.16.	Soil sealing	44
3.5.	Disaggregation of economic data.....	45
4.	HANZE-Events for floods: concepts and contents	47
4.1.	Criteria for inclusion for flood events.....	Error! Bookmark not defined.
4.2.	Database contents.....	Error! Bookmark not defined.
	References.....	49
	Appendix 1. Sources of administrative boundaries for NUTS region map.....	52
	Appendix 2. Detailed categories of non-financial assets	54
	Appendix 3. Local population estimates, 1961-2011.....	54
	Appendix 4. Calibrating urban population change	56
	Appendix 5. Data for the land-use transition model	58
	Appendix 6. Supplementary maps.....	61

Citation

Description papers of HANZE:

Paprotny D., Mengel M. (2023) Population, land use and economic exposure estimates for Europe at 100 m resolution from 1870 to 2020. *Scientific Data*, 10, 372, doi:10.1038/s41597-023-02282-0

Paprotny D., Terefenko P., Śledziowski J. (2023) An improved database of flood impacts in Europe, 1870–2020: HANZE v2.1. *Earth System Science Data*, submitted.

HANZE datasets:

Paprotny D. (2023) HANZE database of historical flood impacts in Europe, 1870-2020. Zenodo, doi:10.5281/zenodo.8221454

Paprotny D. (2023) Pan-European exposure maps and uncertainty estimates from HANZE v2.0 model, 1870-2020. Zenodo, doi:10.5281/zenodo.7885990

This documentation:

Paprotny, D. (2023) HANZE: Historical Analysis of Natural Hazards in Europe – database documentation. Potsdam Institute for Climate Impact Research, Potsdam, Germany.

Description paper of HANZE (version 1 from 2017):

Paprotny, D., Morales-Nápoles, O., and Jonkman, S. N. (2018) HANZE: a pan-European database of exposure to natural hazards and damaging historical floods since 1870. *Earth System Science Data*, 10, 565–581, doi:10.5194/essd-10-565-2018

HANZE dataset (version 1 from 2017):

Paprotny, D., Morales-Nápoles, O., Jonkman, S. N. (2017). HANZE: Historical Analysis of Natural Hazards in Europe. 4TU.Centre for Research Data, Dataset, <https://doi.org/10.4121/collection:HANZE>

Acknowledgements

This update of the HANZE database was prepared with the support of the following project:

“Decomposition of flood losses by environmental and economic drivers” (FloodDrivers), which received funding from the German Research Foundation (DFG), project no. 449175973.

The original HANZE database was prepared with the support of the following projects:

“Risk Analysis of Infrastructure Networks in response to extreme weather” (RAIN), which received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608166.

“Bridging the Gap for Innovations in Disaster resilience” (BRIGAID), which received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 700699.

Disclaimer

The datasets are provided for research purposes only. No warranty is given as to their suitability for user applications. No liability is accepted by the authors for any errors or omissions in the data or associated information and/or documentation.

1. Introduction

HANZE, or **H**istorical **A**nalysis of **N**atural **H**azards in **E**urope, is a database enabling the study of historical trends and driving factors of vulnerability to natural hazards, with a particular focus on floods. It has two components:

- HANZE-Exposure: spatial and tabular data with information on exposed land use, population, production and wealth.
- HANZE-Events: records of past natural disasters, currently limited to floods.

This document presents the characteristics of the database, as well as the methodologies and data sources that were used to construct it.

Important note: HANZE-Exposure was updated and extended in 2022, and all related changes are included in this document. Main differences are highlighted at the end of each subsection (“*Compared with V1*”). An update of HANZE-Events was added in version 2.1 in 2023.

2. Basic characteristics of the database

2.1. Data format

HANZE-Exposure consists of two parts:

- Spatial data, containing raster maps at 100 m resolution.
- Tabular data, containing quantitative data at regional level, together with some additional quantitative and qualitative data at national level.

HANZE-Events consists of tabular data and vector maps.

2.2. Coverage

2.2.1. Spatial coverage

HANZE covers most of the European continent. Included are all 27 European Union member states, all four European Free Trade Agreement members (Iceland, Liechtenstein, Norway and Switzerland), three microstates located in Western Europe (Andorra, Monaco, San Marino), the United Kingdom and one of its Crown Dependencies – Isle of Man. Excluded are, therefore, non-EU successor states of the Soviet Union, as well as Turkey. However, some EU territory is also excluded, namely:

- Canary Islands, Ceuta and Melilla (parts of Spain);
- The Azores and Madeira (parts of Portugal);
- All dependent or overseas territories of EU states, with the exception of the Isle of Man.

Additionally, the following applies to the coverage:

- Data for Cyprus, though exclude areas controlled by the Turkish Republic of Northern Cyprus, cover also the Sovereign Base Areas of Akrotiri and Dhekelia and the United Nations Buffer Zone.
- The Vatican is included, but as part of Rome (region ITI43).
- Kosovo is shown separately from Serbia due to different statistical systems.
- The Isle of Man and Kosovo are often referred here as “countries”, which is done for brevity of the text and without prejudice as to their international status.

The composition of the domain was chosen based on data availability. The domain is shown in Fig. 1.

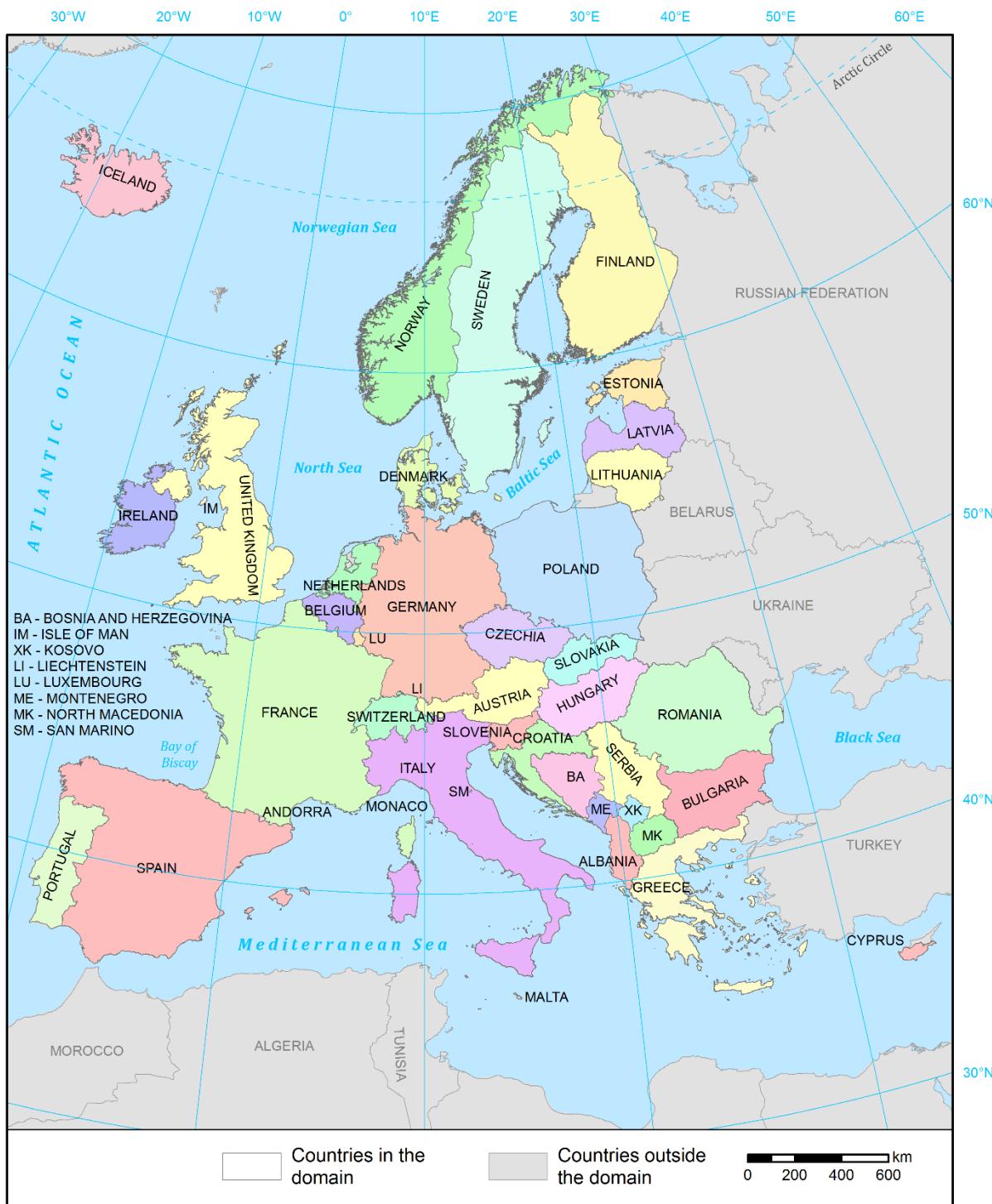


Figure 1. HANZE domain.

Compared to V1: Albania, Bosnia and Herzegovina, Kosovo, Montenegro, North Macedonia and Serbia were added. The Vatican is no longer shown separately, but rather is covered as part of region ITI43 Roma (Rome).

2.2.2. Temporal coverage

HANZE-Exposure covers the period of 1870–2020, with data for 1870–1950 having a 10-year time step, and starting with 1955 a 5-year time step. Additionally, annual data for 2000–2020 are provided. The exception are some economic data (see Table 1), which are given annually for 1870–2020. Stock indicators, such as population numbers, as far as it was possible, refer to 31 December. It should be noted that data for the year 1940 were often compiled from 1938 or 1939 figures due to the disruption caused by World War Two. Such situations are noted in the metadata of the database.

HANZE-Events covers flood events that occurred during the period of 1870–2020, with daily dates of the beginning and end of the events represented in the database.

Compared to V1: Timesteps for 1955 and 1965 were added as well as annual series between 2000 and 2020.

2.2.3. Thematic coverage

As noted in section 2.1, HANZE contains spatial and tabular data. A summary of variables for HANZE-Exposure, together with their spatial and temporal resolution is presented in Table 1. Four variables are the ‘output’ of HANZE, which provides exposure data at 100 m resolution. Additional variables, at regional and national level, are the inputs to HANZE and are used to generate the output variables. Finally, ‘auxiliary’ variables were calculated to support the analysis in HANZE-Events.

Table 1. Variables included in HANZE-Exposure

Category	Variable	Resolution	Reference in document
Output	Land cover/use type	Gridded 100 m, 1/5/10-yearly	3.2.2, 3.3.6, 3.5
Output	Total population	Gridded 100 m, 1/5/10-yearly	3.2.3, 3.3.3, 3.5
Output	Gross domestic product (GDP) per year (euro)	Gridded 100 m, 1/5/10-yearly	3.3.7, 3.5
Output	Value of wealth (euro)	Gridded 100 m, 1/5/10-yearly	3.3.8, 3.5
Output	Soil sealing (%)	Gridded 100 m, 1/5/10-yearly	3.2.4, 3.4.16
Input	Total population	Regions, 1/5/10-yearly	3.3.3
Input	Urban population (% of total population)	Regions, 1/5/10-yearly	3.3.4
Input	Mean number of persons per household	Regions, 1/5/10-yearly	3.3.5
Input	Land use structure, selected types (%)	Regions, 1/5/10-yearly	3.3.6
Input	GDP per year (euro)	Regions, 1/5/10-yearly	3.3.7
Input	GDP structure (% of GDP)	Regions, 1/5/10-yearly	3.3.7
Input	Value of wealth by category (% of GDP)	Countries, 1/5/10-yearly	3.3.8
Input	Efficiency of forest and mining economy	Countries, 2000	3.5
Auxiliary	GDP deflator (base year = 100)	Countries, annual	3.3.9
Auxiliary	Currencies and their conversion factors	Countries, annual	3.3.9

HANZE-Events contains four variables related to flood consequences: area flooded, persons killed, persons affected and monetary value of losses, together with information on date, location and type of event. All variables are described in detail in section 4.2.

2.3. Connection between HANZE-Exposure and HANZE-Events

In order to calculate exposure for a given natural hazard event, the spatial extent of this event needs to be known. Gridded datasets of HANZE-Exposure can be easily intersected with a layer containing extents of events. Here, the extent of past floods was determined as follows. Firstly, in HANZE-Events areas affected by floods were defined using European Union's Nomenclature of Territorial Units for Statistics (NUTS), 2010 edition (see section 3.3.2). Then, 100-year flood zones (river, coastal or combined depending on the type of flood) within those regions were selected from a pan-European dataset developed within RAIN project¹. This allows primarily to:

- 1) Calculate flood losses relative to potential losses during an event.
- 2) Normalize flood losses recorded in different years to a single reference year, i.e. correct for the changes in flood exposure.

An example calculation for the 1953 coastal flood in the Netherlands is presented in Table 2. In the affected regions' 100-year coastal flood zone², according to HANZE-Exposure, population increased by 68% between 1953 and 2020, while GDP increased 5.9 times and wealth 7.3 times. The confidence intervals of exposure are shown, though they are very narrow as the coastal flood zones cover vast majority of the affected NUTS3 regions. It should be noted that because the exposure data are calculated in 5/10-year time steps, the exposure for events that occurred in between the time steps was linearly interpolated.

Table 2. Reported losses, exposure in the potential flood zone of the event, relative and normalized losses for the 1953 coastal flood in the Netherlands. Numbers in brackets are the 95% confidence interval of exposure uncertainty only. Monetary loss is in constant 2020 prices.

Category	Reported losses (1953)	Exposure (1953)	Exposure (2020)	Relative losses (%)	Normalized losses (2020)
Area flooded (km ²)	2000	3917	3917	51.1%	2000
Persons killed (thousands)	1.835	1229 [1208-1265]	2070 [2067-2073]	0.15%	3.09 [3.01-3.14]
Persons affected (thousands)	188	1229 [1208-1265]	2070 [2067-2073]	15.3% [14.9-15.6%]	317 [308-322]
Monetary loss (billion euros)*	5.4	15.0 [14.8-15.2]	87.9 [87.8-88.0]	36.3% [35.7-36.6%]	31.9 [31.4-32.2]
Monetary loss (billion euros)**		60.7 [60.1-61.6]	444.2 [443.9-444.5]	9.0% [8.8-9.1%]	39.8 [39.2-40.1]

* normalized by GDP, ** normalized by wealth.

2.4. List of files

A list of files with explanations is shown in Table 3. All raster and vector files are in ETRS89 / LAEA projection (EPSG:3035). Unless otherwise noted, all raster files have a resolution of 100 m. The table excludes files that can be used to reproduce some of the pre-processing (population disaggregation in section 3.2.4, and probability maps from section 3.4.10)

¹ Data downloadable from Paprotny and Morales-Napoles (2016). For general description of the data, see Groenemeijer et al. (2016), and for details see Paprotny et al. (2017, 2019). For an application of this approach, see Paprotny et al. (2018).

² The affected regions are: Groot-Rijnmond (NL339), Zuidoost-Zuid-Holland (NL33A), Zeeuwsch-Vlaanderen (NL341), Overig Zeeland (NL342), West-Noord-Brabant (NL411).

Table 3. List of files of HANZE database. XXXX = timestep (year).

File	Format	Variables / contents
Output		
CLC_XXXX	8-bit GeoTIFF	Land cover/use type, 44 classes according to Corine Land Cover (section 3.2.2)
Pop_XXXX	16-bit GeoTIFF	Total population per grid cell (in persons)
GDP_XXXX	32-bit GeoTIFF	Gross domestic product (GDP) per grid cell per year (euro in constant 2020 prices)
FA_XXXX	32-bit GeoTIFF	Wealth per grid cell (euro in constant 2020 prices)
Imp_XXXX	8-bit GeoTIFF	Soil sealing degree (%)
Events_floods	Excel file	<i>List of past damaging floods – only available from old repository (see section 4)</i>
Input		
Region_database_population_lu	Excel file	Input land use/cover and population data (list of variables and table structures in section 3.3.1)
Region_database_economy	Excel file	Input and auxiliary economic data (list of variables and table structures in section 3.3.1)
CLC_base_HANZE2	8-bit GeoTIFF	Baseline land cover/use type, 44 classes according to Corine Land Cover
Population_100m	16-bit GeoTIFF	Total baseline (disaggregated) population per 100 m grid cell (in persons)
IMD2012_extent_adjusted	8-bit GeoTIFF	Soil sealing degree in %
NUTS2010_final_100m	Shapefile	NUTS3 (version 2010) region definitions
NUTS2010_100m_c	8-bit GeoTIFF	NUTS3 region definitions (numerical value from vector file attribute table)
gras200a_yld_LAEA2	16-bit GeoTIFF	Agro-climatic potential yield for grass (see 3.4.10)
ylHr0_whe_LAEA2	16-bit GeoTIFF	Output density for wheat (see 3.4.10)
Airports_year_v2	8-bit GeoTIFF	Airports by year of construction
Slope_per_mille_int_masked	16-bit GeoTIFF	Slope (per mille) from EU-DEM
BN_to_urban	16-bit GeoTIFF	Probability map of transition from non-urban to urban after the baseline year
BN_to_crop	16-bit GeoTIFF	Probability map of transition from non-cropland to cropland after the baseline year
BN_to_past	16-bit GeoTIFF	Probability map of transition from non-pasture to pasture after the baseline year
BN_from_crop	16-bit GeoTIFF	Probability map of transition from cropland to non-cropland after the baseline year
BN_from_past	16-bit GeoTIFF	Probability map of transition from pasture to non-pasture after the baseline year

BN_to_crop_p	16-bit GeoTIFF	Probability map of transition from non-cropland to cropland before the baseline year
BN_to_past_p	16-bit GeoTIFF	Probability map of transition from non-pasture to pasture before the baseline year
BN_from_crop_p	16-bit GeoTIFF	Probability map of transition from cropland to non-cropland before the baseline year
BN_from_past_p	16-bit GeoTIFF	Probability map of transition from pasture to non-pasture before the baseline year
ESM2012_street_ext_adj	8-bit GeoTIFF	Surface covered by roads and streets (%) from European Settlement Map 2012
Industry_centroids_int	32-bit GeoTIFF	Distance from centroids of industrial CLC patches in meters
NL_polders	16-bit GeoTIFF	Year of construction of Dutch polders (see section 3.4.15)
CLC_141_142_selected_v2	32-bit GeoTIFF	Green urban areas and sport facilities than are adjacent to selected artificial surfaces (see 3.4.8)
CLC2012_urban_d_int	32-bit GeoTIFF	Euclidean distance from centroids of urban CLC 2012 patches (see 3.4.2)
Clusters2011_high_density_d_int	32-bit GeoTIFF	Euclidean distance from centroids of high-density population clusters (see 3.4.2)
KernelDensityPop1km_int	32-bit GeoTIFF	Kernel population density with 10-km radius (see 3.4.1)
UN_agglomerations_d_int	32-bit GeoTIFF	Euclidean distance from centres of large agglomerations and capital cities (see 3.4.2)
UrbanAudit2018_d_int	32-bit GeoTIFF	Euclidean distance from centroids of cities in Urban Atlas 2018 (see 3.4.2)
VirtualLAU_PD_int_new	32-bit GeoTIFF	Population density of “virtual” LAUs (see 3.4.1)
LAU_data	CSV	Population data (1961-2011) per LAU
Preprocessing data		
Pure_Population_CLC_cells	CSV	GEOSTAT population 1 km grid cells covered with a single CLC class (see 3.2.4)
Population_thresholds	CSV	Thresholds for population disaggregation (see 3.2.4)
ESM2012_buildings	8-bit GeoTIFF	Surface covered by buildings (%) from European Settlement Map 2012
GEOSTAT_extent_adjusted	8-bit GeoTIFF	GEOSTAT 1 km population grid
ESM_GEOSTAT_statistics	CSV	Average population per surface covered by buildings (%)
ESM_Street_GEOSTAT_statistics	CSV	Average population per surface covered by roads and streets (%)
IMP_GEOSTAT_statistics	CSV	Average population per surface covered by impervious surfaces (%)
CLC_changes_sample_data	CSV	CLC land-use transition samples used to train the BN land-use model
CLC_changes_nosample_data	CSV	CLC land-use non-transition samples used to train the BN land-use model

BN_sample_data	NumPy file	Processed CLC land-use transition and non-transition samples used to train the BN land-use model
Validation and analysis data		
BN_sample_data_validation	NumPy file	CLC land-use transition and non-transition samples used to validate the BN land-use model
RAIN_coastalmap_100y	8-bit GeoTIFF	Coastal flood hazard map, 100-year return period from Paprotny et al. (2019)
JRC_flood_mask_100	8-bit GeoTIFF	River flood hazard map, 100-year return period from Alfieri et al. (2014)
LAU2_Austria_pop	Shapefile	Population of Austria by municipality, 1870-2020, for validation
Flood_events_v1.0_list	Excel file	Data on past damaging floods from HANZE v1.0
HANZE-Events		
HANZE_events	CSV (UTF-8)	Flood event data
HANZE_references	CSV (UTF-8)	List of all references
HANZE_events_regions_2010	Shapefile	Flood event data as GIS file (regions v2010)
HANZE_events_regions_2021	Shapefile	Flood event data as GIS file (regions v2021)
Table_S1_countries.csv	CSV (UTF-8)	Country codes/names
Table_S2_regions_v2010	CSV (UTF-8)	Region codes/names, v2010
Table_S3_regions_v2021	CSV (UTF-8)	Region codes/names, v2021
Table_S4_currencies	CSV (UTF-8) CSV (UTF-8)	Data on all currencies used in the study area since 1870
Table_S5_conversion	CSV (UTF-8)	Conversion rates applied to compute losses in 2020 euros
Table_S6_deflators	CSV (UTF-8)	Gross domestic product deflator
Table_S7_deleted_events	CSV (UTF-8)	Flood events in HANZE v1, which were excluded from v2
Regions_v2010_simplified	Shapefile	Region map, v2010, simplified geometry
Regions_v2021_simplified	Shapefile	Region map, v2021, simplified geometry

Compared to V1: all input data needed to run the model are now included. Output map of soil sealing degree was added, while aggregated exposure was excluded, as is the Excel file with past damaging floods, only available from old repository (see section 4).

2.5. Code

The code is stored in the Zenodo repository together with all input data needed to run the code (see section 2.6). It allows to:

- Run the exposure modelling routine to generate all five exposure maps for a single year or multiple years. Alternatively, a hazard zone might be defined and the code then will save the total exposure (population, GDP, fixed assets) per region as a text file. Uncertainty bounds can be calculated in such configuration. In that case the code saves a text file per region and variable (population, GDP, fixed assets) with the 5th, 20th, 50th, 80th and 95th percentile.

- Run the population disaggregation routine (see section 3.2.4).
- Run some of the preprocessing routines: preparing sample data for the Bayesian Network and then generating the BN probability maps (see section 3.4.10); computing population thresholds for the disaggregation routine (see section 3.2.4).
- Run the validation calculation, which allows: comparing the modelled population with observation for Europe (1960-2010) and Austria (1870-2020); comparing land-use change with observations in Corine Land Cover (see Appendix 5).
- Visualizing exposure for past floods (from HANZE v1.0) in the form of graphs (population trends with uncertainty) and maps (land cover/use, population and fixed asset distribution in 1870, year of the event and 2020).
- Visualizing exposure for individual NUTS3 regions as maps (land cover/use, population and fixed asset distribution in three timesteps, default 1870, 1950, 2020).

The user can use the exposure model outside the defined spatial extent, regional definitions or years by changing the input data, as follows:

- To run different years that included in the database, the user needs to add data for the additional years to two files: “Region_database_population_lu.xlsx” and “Region_database_economy.xlsx”.
- To run a different NUTS region definition, the data needs to be added to the two Excel files mentioned above, as well as replacing “NUTS2010_final_100m.shp” with a shapefile with the new definitions, but with the same type of data in the columns, and “NUTS2010_100m_c.tif” containing numerical values from the “gridcode” column of the shapefile. The spatial extent of the raster must be exactly the same as other rasters.
- To run the model in a different area, all raster files need to be adapted so that they have the same spatial extent.

Additional code enables reproducing the analysis of HANZE v2.1 flood impact data in the ESSD paper:

- Generating a GIS file from tabular impact data;
- Aggregating the number of events per NUTS 3 region.

Compared to V1: code of the model is now publicly available for the first time.

2.6. Data storage

The data and code are stored as follows:

- Code and this documentation: <https://dx.doi.org/10.5281/zenodo.8223833>
- Input data: <https://dx.doi.org/10.5281/zenodo.6783023>
- Output data: <https://dx.doi.org/10.5281/zenodo.7885990>
- Auxiliary building asset value dataset (see section 3.3.8):
<https://dx.doi.org/10.5281/10.5281/zenodo.6573503>
- Flood event data: <https://dx.doi.org/10.5281/zenodo.8221454>
- HANZE v1.0: <https://doi.org/10.4121/collection:HANZE>

3. HANZE-Exposure: methodology and detailed contents

3.1. Outline

The general concept of the methodology is based on HYDE database from the Netherlands Environmental Assessment Agency (Klein Goldewijk et al. 2010, 2011). Firstly, two detailed maps of population and land use are compiled for a single time point. Complete surveys of those variables with a high spatial resolution are very few, and datasets constructed with a certain methodology rarely extend beyond a single time point. Therefore, once the two maps are collected—we can dub them ‘baseline maps’—other time points in the past and in the future could be calculated based on the baseline maps. In this study, the baseline maps refer to the year 2011/12, and have a spatial resolution of 100 m. For the years between 1870–2020 only know the total population and land use at NUTS 3 regional level is known. Hence, for each time step, the population and the different land use classes had to be redistributed inside each NUTS 3 region in order to match the regional totals. Several methodologies were used in order to provide the best approximation of spatial distribution of each land use class and population. Efforts were concentrated on estimating past and future residential urban areas (where most population is settled) and lands used by agriculture and infrastructure.

The procedure is summarized in Fig. 2; the primary input for the model are baseline maps (see section 3.2) and a database of regional- and country-level statistical data (section 3.3). The model of changes in land use and population distribution (section 3.4) generates population/land use maps which are used to disaggregate economic data (section 3.5) to produce final exposure maps with a 100 m resolution (3.6).

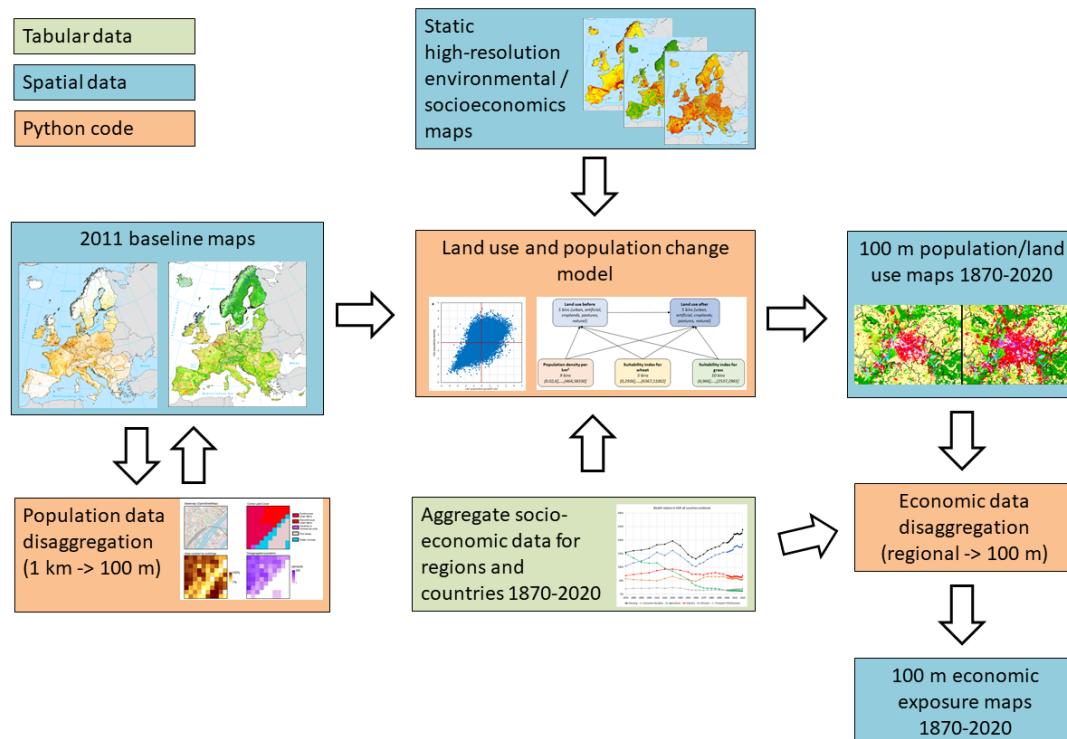


Figure 2. Workflow of HANZE-Exposure.

3.2. Baseline maps

3.2.1. Administrative boundaries

A single administrative boundaries map was used in the database to collect statistical data and carry out land-use modelling. The European NUTS classification, version 2010 was used (see section 3.3.2). The semi-official map of NUTS regions available through Eurostat has relatively low resolution (1:1,000,000) and does not have an open licence, hence a new high-resolution map was prepared for this study by combining OpenStreetMap (2021) administrative boundaries (usually through aggregation of municipalities or similar units) with open data from national agencies for certain countries. OpenStreetMap data was extracted through QGIS with ‘QuickOSM’ plugin. The other datasets used are listed in Appendix 1. Where necessary, manual adjustments to administrative divisions were made due to changes between NUTS version 2010 and available boundary layers. The combined layer for Europe was adjusted to fit the coastline indicated in Corine Land Cover 2012 (see section 3.2.2) and converted into a 100-meter resolution grid. With those transformations, full consistency with the baseline land cover/use layer was achieved.

3.2.2. Land cover/use

The baseline land cover/use is based on Corine Land Cover (CLC) 2012, version 20u1 (Copernicus 2021). CLC is a project initiated in 1985 and supervised by the European Environment Agency. It has since produced four pan-European land use maps for 1990, 2000, 2006, 2012 and 2018. The maps are prepared, in general, by manual classification of land cover patches from satellite imagery. For the 2012 edition, images collected during 2011-2012 were used. The inventory consists of 44 classes. The minimum size of areal phenomena is 25 hectares. For linear features (roads, railways, rivers etc.), a minimum width of 100 m is used. The thematic accuracy of the dataset is, according to product design specifications, above 85%, which was generally achieved according to the validation report (Copernicus 2017). It should be also noted that mapping is done by each country independently, and therefore the classification of land use is not always fully consistent between countries. For instance, a complete lack of ‘continuous urban fabric’ class is noticeable over the Netherlands, despite this class being typically used for downtown areas of larger cities in all other countries.

CLC 2012 covers the entire domain with one exception: Andorra. For this country the map was constructed by overlaying data from 4 different sources, in order:

- 1) CLC 2012 v20u1, which covers a small strip around the border;
- 2) CLC 2000 v20u1, which covers a larger strip around the border;
- 3) Open Street Map vectors obtained in 2016 from Gisgraphy (2019);
- 4) Global Land Cover 2000 (Joint Research Centre 2015).

The final map for the study area is presented in Fig. 3, with all CLC classes shown in Fig. 4.

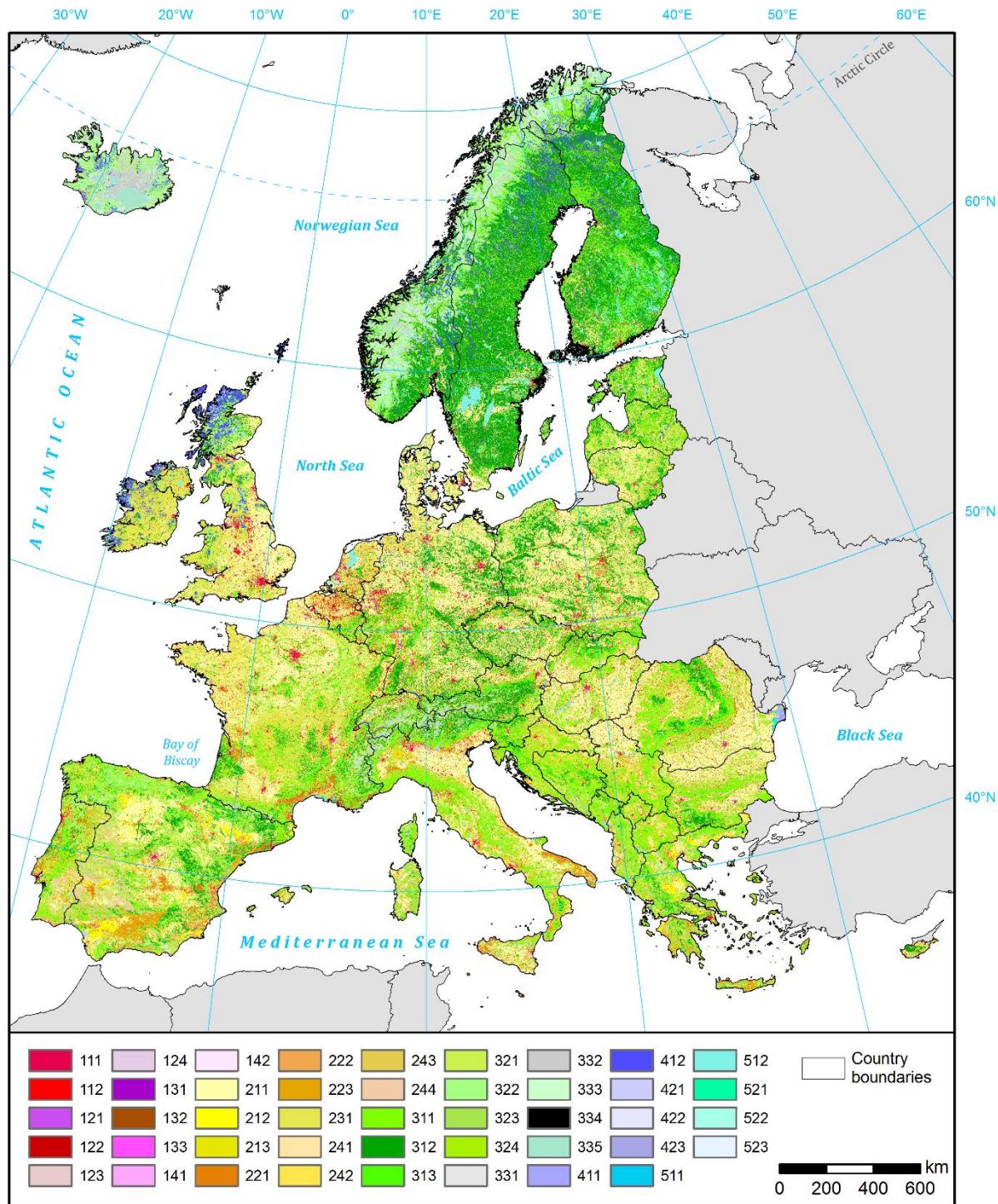


Figure 3. Baseline land cover, 2011, based on Corine Land Cover 2012. For explanation of CLC classes, see Fig. 4.



Figure 4. Corine Land Cover classes. Source: European Environment Agency (2011).

Compared to V1: The baseline map was recompiled using a new version of the CLC inventory (20u1 from 2020 compared with 18_5_1 from 2016).

3.2.3. Soil sealing

Impervious surfaces are the artificial substitutions of natural land cover, which create an impermeable cover of soil. This has a significant impact on hydrological properties of a given area and, consequently, on flood frequency and intensity. The degree of imperviousness (also called soil sealing) is one of the input parameters of LISFLOOD model used to generate pan-European flood maps (Alfieri et al. 2014). HANZE-Exposure estimates change in soil sealing by modifying the baseline soil sealing map. The map is the Imperviousness Density 2012 dataset from Copernicus (2021). It was created by algorithmic classification of high-resolution satellite images³ with a calibrated normalised difference vegetation index (NDVI). The native resolution of the dataset is 20 m, but it was aggregated to 100 m for

³ mainly SPOT 5 and 6 missions, with a resolution of 2.5 m.

consistency with the land cover map. The thematic accuracy is, according to product design specifications, above 90%, though the validation report from Copernicus (2019) indicates that this is not always achieved.

3.2.4. Population

The baseline population map is based on GEOSTAT population grid for the year 2011, version 2.0.1 (Eurostat 2021b). This dataset has a 1 km resolution and for most countries it represents the actual population enumerated and georeferenced during 2011. However, for some smaller countries, namely Andorra, Bosnia and Herzegovina, Cyprus, Iceland, Isle of Man, Luxembourg, Monaco, Montenegro, North Macedonia, San Marino, Serbia and the Vatican, the data are estimates provided by the Joint Research Centre. In those cases, the census data were disaggregated from census enumeration blocks or local administrative units to a 1 km grid based on land use data from a refined version of Corine Land Cover 2006. In case of Finland and Sweden, the population grid was made for a different date than the census. Detailed information per country can be found in the population input dataset. The GEOSTAT dataset is shown in Fig. 5.

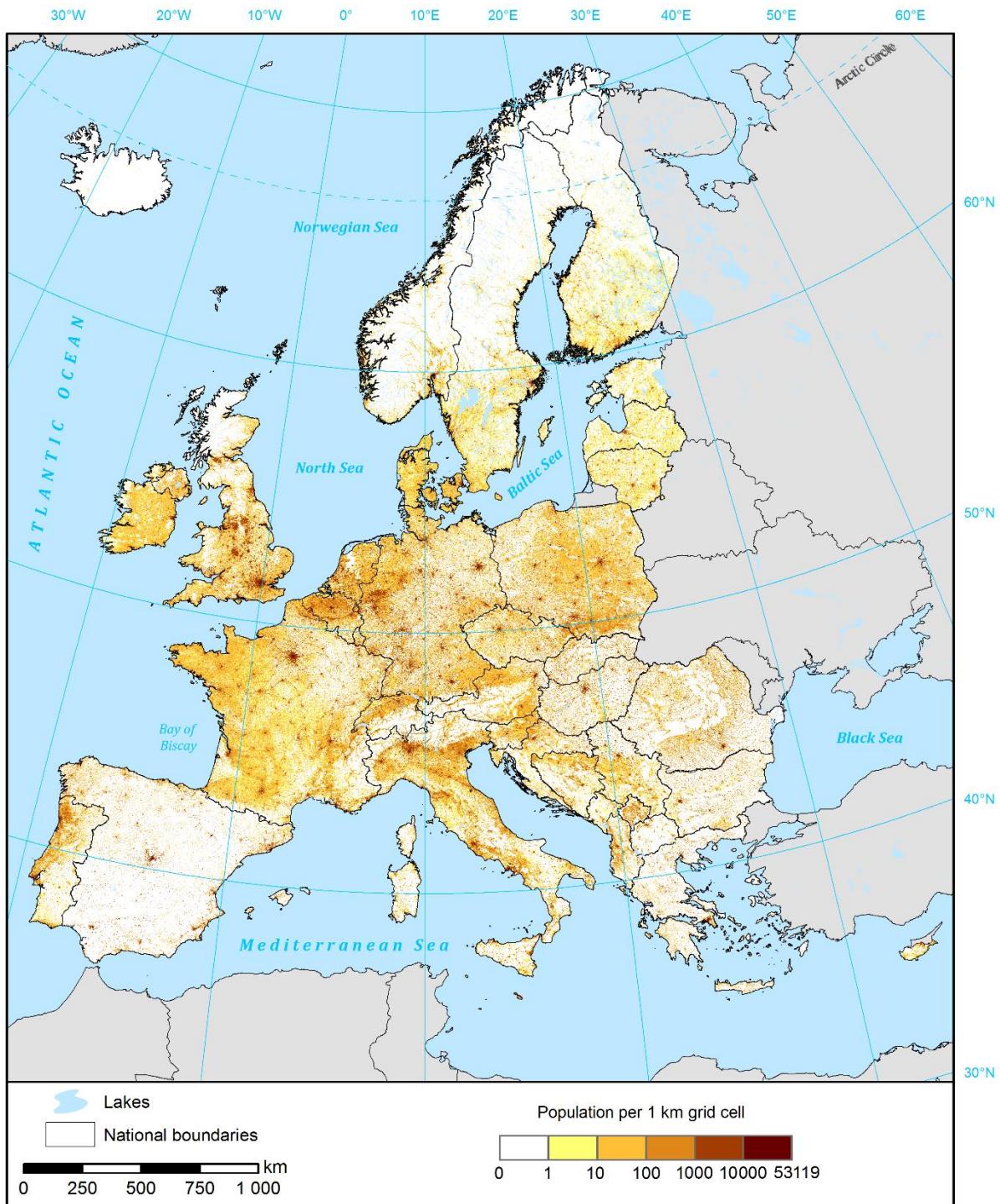


Figure 5. GEOSTAT population grid, 1 km resolution.

For this study, the 1 km grid had to be further disaggregated to a 100 m resolution. Several methods have been proposed for this procedure, and also tested for Europe (Gallego 2010, Gallego et al. 2011, Batista e Silva et al. 2013). Here, we combine methods M1 and M3 described in Batista e Silva et al. (2013). M1 denotes the ‘limiting variable method’ used in cartography for creating dasymetric maps of population density at least since the 1930s (Wright 1936). The procedure is an iterative algorithm applied separately for each 1 km grid cell. This procedure is as follows:

- Firstly, uniform population density is assigned for each land use class in a 1 km grid cell:

$$Y_{LG}^0 = Y_G = \frac{X_G}{S_G} \quad (1)$$

where Y_{LG}^0 is the population density for land use $L \in \{1, \dots, n\}$ in grid cell G at step 0, Y_G is the population density in the grid cell, i. e. population number X_G divided by area S_G .

- A population density threshold T_L is defined for each one of n land use classes.
- Land use classes are ranked and the subindex L is renumbered from lowest to highest population density, i.e. $L = 1$ denotes the least densely populated land use class in the grid cell
- Proceeding in order starting with $L = 1$, in step L the density attributed to class L in the previous step is modified if it is above the threshold, i.e. if $Y_{LG}^{L-1} > T_L$. That creates a surplus population U_{LG}^L :

$$U_{LG}^L = S_{LG} \times (Y_{LG}^{L-1} - T_L) \quad (2)$$

- Surplus is then redistributed among the remaining land use classes M , hence:

$$Y_{LG}^L = T_L \quad (3)$$

$$Y_{MG}^L = Y_{MG}^{L-1} + \frac{U_{LG}^L}{\sum S_{MG}}, M > L \quad (4)$$

- If after completing all iterations there is still surplus population, i.e. if $X_G > \sum T_L S_{LG}$, it is redistributed proportionally to the threshold:

$$Y_{LG} = \frac{T_L X_G}{\sum T_L S_{LG}} \quad (5)$$

The crucial aspect of this method is defining the thresholds T_L . Here, we use thresholds as suggested by Eicher and Brewer (2001), i.e. the 70th percentile of the population density of grid cells for which only one land use class was reported in our baseline land use map. Such “pure” cells constituted around 5% of all population grid cells. Gallego et al. (2011) have shown that a different definition of thresholds works slightly better for Europe; however, the authors used population data by communes, which are not used here, and which their method would require in combination with gridded data. The final thresholds T_L are shown in Table 4. For artificial surfaces other than urban fabric, the CLC classes were merged for the threshold calculation, as very few, if any, “pure” cells could be found for each of those classes. Also, for all areas covered by wetlands, water, sand, glaciers, bare rocks or burnt vegetation the threshold was set at 0, as those terrains are in principle uninhabitable.

As an additional limitation, only those land use classes in a given cell were used, which contained any man-made structures of particular kind. Three remote-sensing gridded datasets (100 m resolution) were used here; if no land use class in a cell possessed any structures from the first dataset, the second was used, then third if necessary, as follows:

1. Buildings;
2. Impervious surfaces;
3. Roads and streets.

Buildings and streets were obtained from European Settlement Map 2012 (Release 2017) and impervious surfaces from Imperviousness Density 2012, both from Copernicus (2021). If no structures were present in the 1 km cell (as they were not detected in the satellite images), all land use classes were utilized.

Table 4. Thresholds for population disaggregation algorithm T_L

CLC class name and code	Threshold (persons per km ²)
Continuous urban fabric (111)	22110
Discontinuous urban fabric (112)	6431
Industrial or commercial units (121–142)	90
<i>Other artificial</i> (122–142)	31
Non-irrigated arable land (211)	31
Permanently irrigated land (212)	52
Rice fields (213)	10
Vineyards (221)	47
Fruit trees and berry plantations (222)	42
Olive groves (223)	56
Pastures (231)	40
Annual crops associated with permanent crops (241)	61
Complex cultivation patterns (242)	79
<i>Land principally occupied by agriculture</i> (243)	48
Agro-forestry areas (244)	7
Broad-leaved forest (311)	9
Coniferous forest (312)	6
Mixed forest (313)	8
Natural grasslands (321)	13
Moors and heathland (322)	13
Sclerophyllous vegetation (323)	7
Transitional woodland-shrub (324)	12
Sparsely vegetated areas (333)	19
<i>Uninhabitable natural areas</i> (331–332, 334–523)	0

The result of the calculation, however, is only the population per land use L in each 1 km grid cell G . Hence, the population had to be disaggregated further, and for that we used an approach similar to method M3. This method redistributes the population proportionally to the density of man-made structures. This variable has a range from 0%, which indicates completely natural surface, and 100%, which indicates land completely sealed by an artificial surface. The three datasets were used, primarily buildings from the European Settlement Map (ESM) 2012. If no buildings were indicated in a 1 km cell, imperviousness was used instead. In case no soil sealing was detected, roads and streets from ESM 2012 are used. This can happen mainly because ESM 2012 combined remote sensing data with multiple other sources (e.g. OpenStreetMap, European Union's Urban Atlas and Tom Tom's Tele Atlas), while Imperviousness Density 2012 is entirely a remote-sensing based product.

ESM and Imperviousness datasets have very high native resolutions (2.5 and 20 m, respectively). A resampled a 100-meter resolution version provided by Copernicus (2021) was used for the disaggregation, while for computing the dependency between surface density and population was determined using data further resampled to a 1 km grid. In the process, average population density in grid cells with given artificial surface density could be calculated. The resulting dependencies can be approximated as power functions (Fig. 6). Very few cells had very high average % of surface covered by structures, hence the functions were computed from values ranging from 1% to 16% (roads and streets), 64% (buildings) and 84% (impervious surfaces). Hence, the population X_g in 100-meter grid cell g is equal to:

$$X_g = \frac{Z_g}{\sum Z_g} Y_{LG} S_{LG} \quad (6)$$

where Z_g is the population of grid cell g obtained from the power function divided by maximum population:

$$Z_g = \frac{BV_g^A}{8000} \quad (7)$$

where V_g is the imperviousness in grid cell g . The maximum population was defined as 8000 as all three datasets reached peak population density around this value. The parameters A and B are indicated in Table 5.

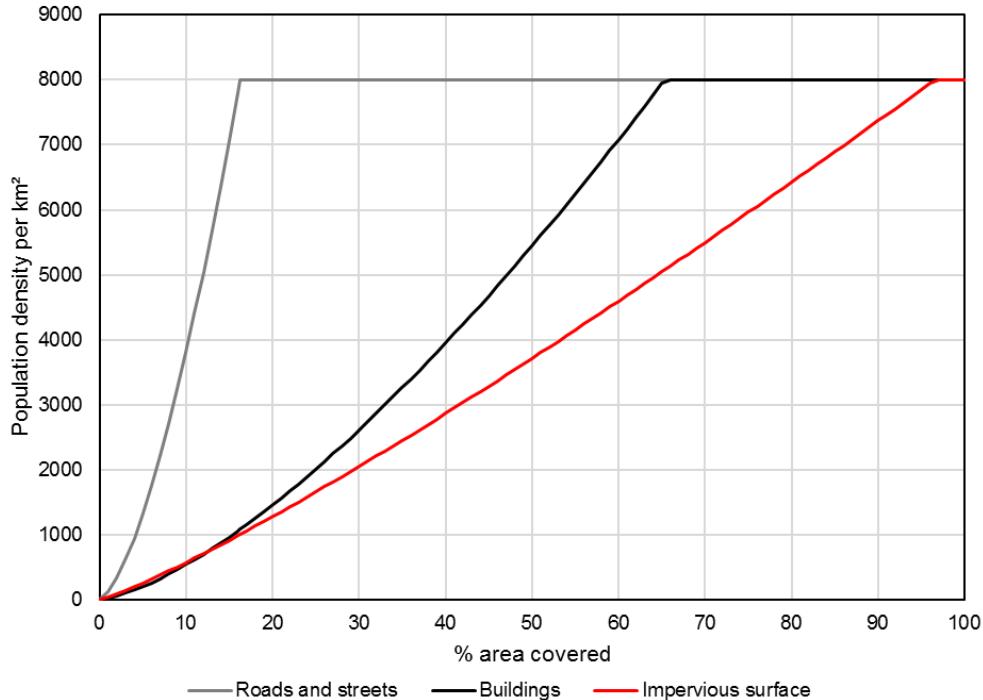


Figure 6. Dependency between average population density per % area covered used for population disaggregation.

Table 5. Parameters A and B in eq. 7

Dataset	A	B
Buildings	1.4399	19.4875
Impervious surfaces	1.5244	113.9358
Roads and streets	1.1658	38.8367

The population X_g is rounded, as population numbers need to be integers. Consequently, the population was added or subtracted by iteratively reducing population numbers in 100-meter cells starting with cells in which the smallest change in unrounded value would change the rounded value. In some cases, more than one 100-meter cell had equal values and the 1-km population couldn't be matched. Then, population was added or subtracted by iteratively reducing population numbers by 1 at a time starting with 100-meter cells with the highest population. If again there were cases of multiple cells of equal values, 100-meter cells with higher % of area covered by structures were used.

If no data was available or the % values were the same, the population is added or subtracted randomly within the equal cells.

An example of the disaggregation is shown in Fig. 7. The area shown corresponds to a 1 km grid in the GEOSTAT population dataset over the city of Szczecin, Poland. In this grid cell, the population at the time of the 2011 census was 5230. The top left box is an extract from OpenStreetMap. The top right box shows the land use structure according to Corine Land Cover 2012, and the bottom right box shows the percentage of area covered by building according to ESM 2012 dataset. The final 100 m population grid, based on aforementioned disaggregation process, is presented in the bottom left box.

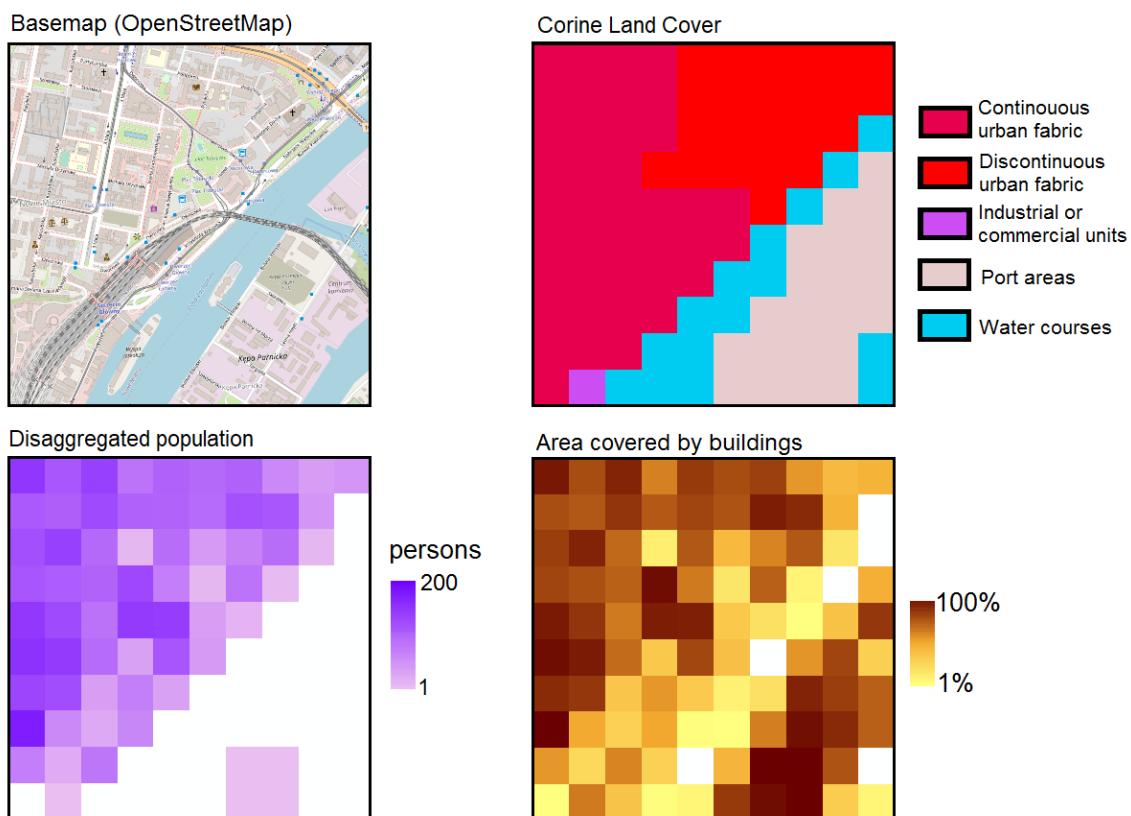


Figure 7. Disaggregation result and source data (population in the grid = 5230), contrasted with OpenStreetMap basemap. Fragment of the city centre of Szczecin, Poland (region PL424).

Compared to V1.0: the population thresholds were revised and the detailed allocation of population is now primarily based on building cover from ESM 2012 and optionally from imperviousness and roads/streets from ESM 2012 rather than only based on imperviousness. Also, only CLC classes with some man-made construction are used in the disaggregation process in a given 1-km grid cell, not all classes available. The mechanism of rounding the disaggregated population and matching the desired total population in the 1-km cell was updated to reduce the random component. A new example was produced.

3.3. Input database of historical statistics

3.3.1. Structure of input dataset files

In this section the structures of two input datasets (in *.xls format) are described (Tables 6 and

7).

Population and land cover/use - Expo_input_CLC_Pop**Table 6. Contents of population and land cover/use data file - Expo_input_CLC_Pop**

Variable	Unit	Table structure
Population	Thousands of persons	Code – NUTS3 region code Name – NUTS3 region name 1870...2020 – data by year
Urban fraction	Urban population as % of total population	
Persons per household	Mean number of persons	
Croplands	% of total area	
Pastures	% of total area	
Forests	% of total area	
Infrastructure	Area covered by road and rail infrastructure in ha	
Census information	Additional information on the 2011 censuses, which are the baseline population figures	Code – NUTS0 country code Name – country/territory name Date – census date Type – census type Source – method of collecting population data GEOSTAT accuracy – information on gridded data production methods
Airports	Airports identified in the CLC data	CLC – Corine Land Cover 2012/2018 vector polygon code Name – airport name Year – year of construction NUTS3 – NUTS3 region code ICAO – airport ICAO code IATA – airport IATA code
Reservoirs	Reservoirs identified in the CLC data	CLC – Corine Land Cover 2012/2018 vector polygon code Name – name of dam Year – year of construction NUTS3 – NUTS3 region code GRAND – reservoir code in GRaND database

Economy - Expo_input_Econ**Table 7. Contents of economic data file - Expo_input_Econ**

Variable	Unit	Table structure
GDP	Million euro in constant 2020 prices	Code – NUTS3 region code Name – NUTS3 region name 1870...2020 – data by year
GDP from agriculture	% of total GDP	
GDP from industry	% of total GDP	
Wealth in housing	% of total GDP	
Wealth in agriculture	% of GDP from agriculture	Code – NUTS0 country code Name – country/territory name 1870...2020 – data by year
Wealth in industry	% of GDP from industry	

Wealth in services	% of GDP from services	
Wealth in infrastructure	% of total GDP	
Wealth in consumer durables	% of total GDP	
Forestry index	Forest economy efficiency relative to agriculture (%)	Code – NUTS0 country code Name – country/territory name Index – relative efficiency for year 2000 (%)
Mining index	Mining economy efficiency relative to industry (%)	Code – NUTS0 country code Name – country/territory name Index – relative efficiency for year 2000 (%)
GDP deflator	Index, base year = 100	Code – NUTS0 country code Name – country/territory name 1870...2020 – data by year Unit – unit of measure (2020 = 100, 1990 = 100 or 1913 = 100)
Currencies	List of all currencies used in the period	See Table 7, section 3.3.9
Currency conversion	Conversion factors to euro (euro = 1). For countries not currently using euro, 2020 exchange rates were used.	Country – NUTS0 country code Currency – currency code Code – merged NUTS0 and currency code Conversion – conversion factor

Both files also contain:

- Sources, which explain the origin of data, transformations made to the original data and methods to estimate gaps in the data, divergence from standard definitions and other relevant information on the data.
- References, which lists all publications mentioned in “Sources”.

Compared to V1: (1) data for six countries were added; (2) wealth in consumer durables was added; (3) forestry index was changed and mining index was added; (4) reference year was changed from 2011 to 2020; (5) large update and revision of existing data, with the number of sources cited increased from 271 to 375 (116 added, 12 removed).

3.3.2. NUTS 3 regions

The regional boundaries are taken from European Union’s Nomenclature of Territorial Units for Statistics (NUTS). This classification has 4 levels (0, 1, 2, 3), where 0 is the national level and 3 is the finest regional division. The 2010 version of NUTS is used here (European Union 2011), which was used for dissemination of statistics during 2012–2014, including 2011 population and housing census data. NUTS favours administrative divisions in defining the regions, though often statistical (analytical) regions are used instead, by amalgamating smaller administrative units. The goal is to obtain, at a given level, regions that have similar number of inhabitants. For example, the regions in the Netherlands are defined as follows:

- NUTS 1: 4 statistical regions (*Landsdelen*);
- NUTS 2: 12 provinces (*Provincies*);
- NUTS 3: 40 statistical regions (*COROP-gebieden*).

It can be noticed that only at NUTS 2 level the actual administrative divisions of the Netherlands are used, while the NUTS 1 and 3 regions are groups of provinces and municipalities, respectively. Some small territories have no NUTS subdivisions (Andorra, Cyprus, Liechtenstein, Luxembourg, Montenegro, Monaco, San Marino), while Kosovo and Bosnia and Herzegovina are not covered by the

NUTS system. Therefore, subdivisions based on administrative regions of those countries were defined and coded in a manner consistent with the NUTS system.

In the database there is a total of 1422 regions defined (Fig. 8). A region has an average area of 3550 km² and an average total population of 373,000 as of 2011 census. A bit more than a quarter of all regions are located in Germany (412), since they are typically smaller than in most other countries (average population is only 195,000).

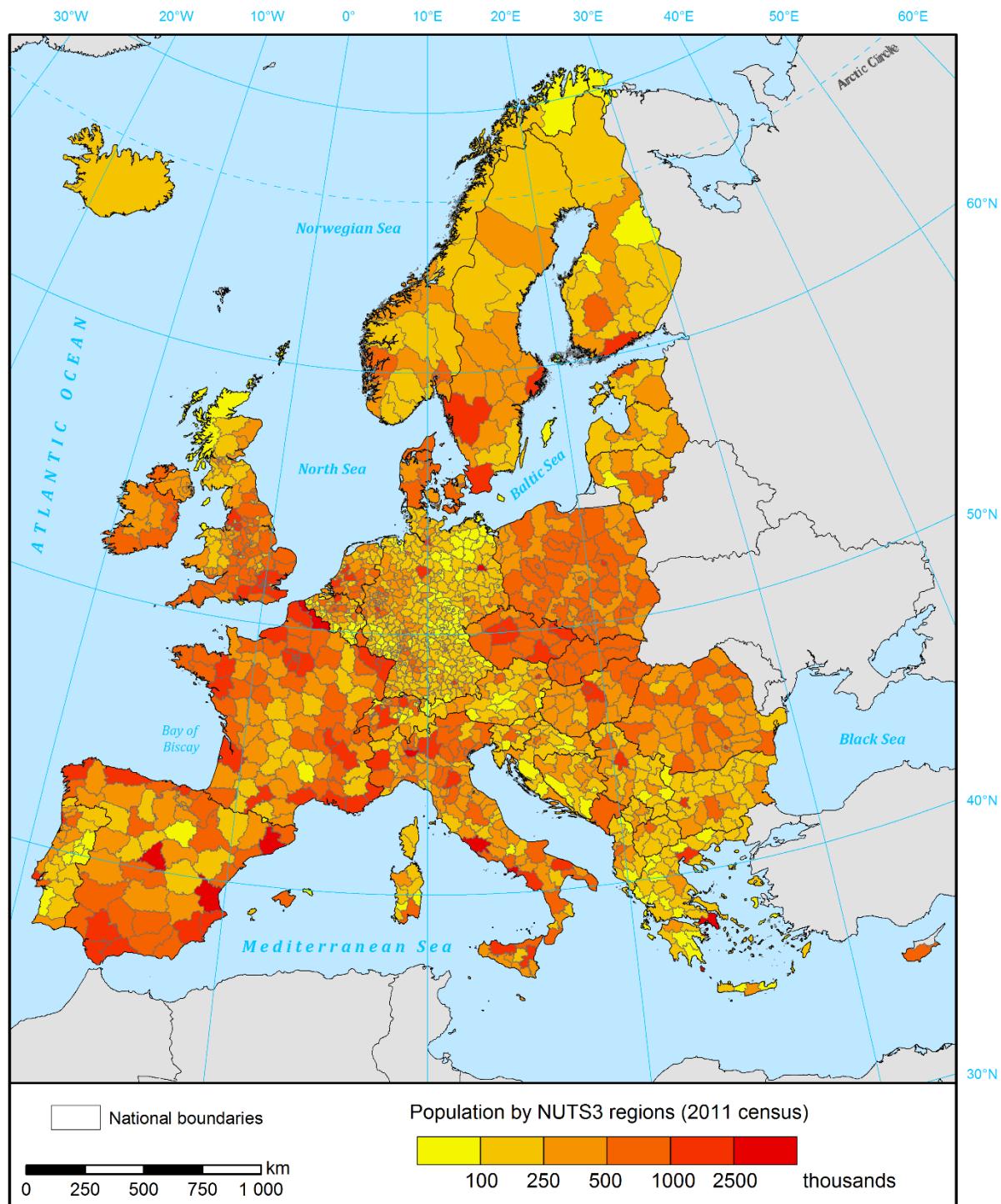


Figure 8. NUTS 3 regions (2010 version) in the study area and their population.

Compared to V1: Regions for new countries in the study were added and a completely new map of regions was compiled from open data (see section 3.2.1), replacing a less accurate map used before.

3.3.3. Total population

Total population refers to the overall number of persons living in a region. Population can be defined as *de facto* population, i. e. the number of persons physically present in an area at a given moment of time, or *de jure*, i.e. the number of persons usually resident in an area, excluding short-term movements or migrations of population (United Nations 2015)⁴. The prime sources of population figures are censuses, held typically every decade, supplement by annual balances of births, deaths and migrations. Starting with the 1970s, many European countries gradually replaced censuses with population registers, providing continuous information on population size. For this database, statistics were generally compiled from country-specific sources, though for 1960-2020 data from Eurostat (2021a) were mostly used, which included recalculation of historical census data to modern administrative divisions, and annual population estimates starting with 1990. For some non-EU countries with no subdivisions into regions from 2019 estimate and projections by United Nations (2019) were also used.

However, data at current administrative divisions were not always available. In several cases, historical divisions were recalculated using one of two methods: the ‘population method’ and the ‘territorial method’. The ‘population method’ recalculates the population of ‘old’ administrative divisions to ‘new’ ones by shifting overlapping proportions of population between the territorial units. More formally, the population X_A^t in each post-reform (‘new’) administrative unit A in year t is a sum of fractions F_{AB} multiplied by the population X_B^t of pre-reform units B:

$$X_A^t = \sum F_{AB} X_B^t \quad (8)$$

The fractions F_{AB} could only be determined if both populations X_A^t and X_B^t are known for the same year; in other words, F_{AB} is the percentage of B’s population living within the boundaries of A. Yet, the extent of administrative changes may not allow to calculate the fractions. The ‘territorial method’, on the other hand, requires a digital map of both pre- and post-reform administrative divisions. The fraction F_{AB} is then the percentage of B’s territory also belonging to A. This assumes equal population density within A and B, therefore this information could only be used to determine population growth rates X_A^t/X_A^{t-1} . Those growth rates were used to extrapolate the population from the earliest year for which data for A are known. It should be noted that both methods could be used for different time periods for the same country, also multiple times, in order to achieve population estimates for the 2010 version of NUTS 3 regions. The two methods were used depending on the availability of data.

3.3.4. Urban population

The fraction of the overall number of persons living in a region that reside in areas defined as urban. The definitions of urban areas vary from country to country (United Nations 2015); the criterion could be administrative (legally designated cities or towns), demographic (all settlements or communes with more inhabitants than a given threshold) or statistical, based on multiple criteria (population number or density, percent of non-agricultural employment, distance between buildings etc.). For the purpose

⁴ Countries typically have their own, specific rules what counts into their population figures, deviating to a various degree from the *de facto* and *de jure* concepts. Such differences are mostly not relevant relative to countries’ overall population size.

of this study, the urban population is defined as the population disaggregated into CLC classes 111 and 112 (urban fabric); the remainder of the population is therefore considered rural.

However, the disaggregation procedure (section 3.2.3) was only done for the 2011 baseline map. Therefore, national definitions of urban populations were used to determine growth rates of urban and rural populations, which were used to extrapolate the urban fraction from the baseline map. For some countries, different definitions were used in different time periods. They could all be used, however, as the various time series overlapped, allowing them to be linked to the 2011 map. The data was mostly collected from national sources, supplemented by United Nations (2014) and other international yearbooks.

3.3.5. Mean number of persons per household

The total population divided by the total number of households in a region. Typically, a household is defined as one or more people who occupy a single housing unit (Haupt et al. 2011). Households consists both of private households and collective households, i.e. institutions such as prisons, nursing homes, dormitories, homeless shelters, army barracks etc. (United Nations 1974). However, the statistics on the latter were not always available, though this has minor effect on the accuracy of the estimates, as population in institutions typically do not exceed 1% of total population (according to data published in United Nations yearbooks). Additionally, data on the number of dwellings was sometimes used if the number of private households was not available. Usually the difference between the two statistics is negligible (some dwellings may not be occupied, while some might contain more than one household). The data was mostly collected from national sources, supplemented by several international compendia.

3.3.6. Land use structure

The region's area, or its percentage, covered by different land use classes. The definitions vary between countries; for the purpose of this study, the 2012 statistics were obtained directly from the baseline land use map (section 3.2.2). The following land use classes were calculated:

- Croplands: CLC classes 211-213 “Arable land”, 221-223 “Permanent crops” and 241-244 “Heterogeneous agricultural areas”;
- Pastures: CLC class 231 “Pastures”;
- Forests: CLC classes 311-313 “Forests”;
- Infrastructure: CLC class 122 “Road and rail networks and associated land”.

For years 2000-2018, the data were obtained or interpolated from Corine Land Cover datasets, and the trend in land use change was extrapolated to 2020. For 1870-1995, area covered by croplands, pastures and forests was extrapolated using different data series following various definitions. For more recent years, regional data from Eurostat (2021a) were largely used, otherwise national statistics and FAO (2021a), including old FAO forest yearbooks, were used after 1940. Before that, if no national statistics or research monographies were available, HYDE 3.2 (Klein Goldewijk 2010, 2011) database was used to extend the series for croplands and pastures back to 1870. HILDA dataset (Fuchs et al. 2013, 2015) was used to extend the data for forests back to 1900, and in a few cases also croplands and pastures back to 1900.

Area covered by road and rail infrastructure was extrapolated using statistics on motorway and railway length, mostly from national statistics, Eurostat (2021a) and Mitchell (1998).

Compared to V1: Database of forest share is now complete and used in the analysis.

3.3.7. GDP and its composition

GDP is the gross domestic product, i.e. value of an economy's total output of goods and services, less intermediate consumption, plus net taxes on products and imports, in a specified period (European Union 2013). Here, we include estimates of GDP at constant prices, adjusted to 2020 price levels, with average currency exchange rates in 2020 used to convert GDP value to euro. The starting point for all countries, except for the microstates, are Eurostat's GDP data at regional level calculated using the 2010 European System of National and Regional Accounts, or ESA 2010 (European Union 2013). GDP was calculated in the past with a variety of methodologies, while for the early 20th and late 19th centuries GDP estimates are often based on proxies. Therefore, the different time series of data were linked to current Eurostat estimates. However, the timeseries were checked if the GDP per capita was always above the subsistence level, which is the lowest possible GDP per capita for sustained periods of time (Bolt et al. 2018). The subsistence level was assumed to be the World Bank poverty line (1.90 USD per day) adjusted to 2020 euros and then to purchasing power parities in each country (from Eurostat database). For 59 regions, mostly in Bulgaria and Romania, GDP per capita was increased to the subsistence level in the late 19th century.

Data on GDP by sector were also collected. Strictly, they represent the percentage composition of gross value added (GVA), a subcomponent of GDP (GDP minus net taxes), as data on net taxes are not collected by sector. Nevertheless, the GVA composition was applied to GDP. The following sectors were distinguished, based on NACE Rev. 2 (European Union 2013):

- Agriculture: Agriculture, forestry and fishing (A);
- Industry: Mining and quarrying (B), Manufacturing (C), Electricity, gas, steam and air conditioning supply (D), Water supply; sewerage, waste management and remediation activities (E);
- Services: construction (F) and all remaining sectors (G-U).

As can be noticed, the difference between traditional three-sector split is the inclusion of construction in services rather than in industry. The data sources, apart from Eurostat and some international compilations, were mostly country-specific.

Compared to V1: the GDP per capita timeseries are now checked if they are at least at the subsistence level.

3.3.8. Wealth and its composition

“Wealth” is considered here in a narrow sense, and relates to assets that could be destroyed during a natural hazard and conceivably contribute to reported losses. Therefore, “wealth” is comprised of tangible fixed assets. Fixed assets are produced non-financial assets that are used repeatedly or continuously in production processes for more than one year. They consist of dwellings, other (non-residential) buildings and structures, machinery and equipment, and cultivated biological resources. Therefore, the following items are excluded: all financial assets, intangible assets (e.g. patents and software), inventories of produced goods, valuables, natural resources (incl. land, subsoil assets and non-cultivated biological resources) and consumer durables (European Union 2013). More detailed classification of fixed assets is shown in Appendix 1.

Statistics on tangible fixed assets according to ESA 2010 methodology are available from Eurostat for most, though not all, countries. However, the Eurostat series mostly start in 1995, and were amended with OECD (2021), Goldsmith (1985) and several other compilations and country-specific sources. Historical series were linked to Eurostat's ESA 2010 estimates, where available. Consumer durables for

recent years were computed using methodology outlined by Paprotny et al. (2020), with the addition of personal vehicles⁵ to the list of durables covered by that study, as this type of asset is not routinely computed by statistical offices. The data was extrapolated into the past using Goldsmith (1985), Piketty and Zucman (2014) or country-specific sources.

The value of assets is measured in current replacement costs, i.e. the market or basic cost of replacing an asset in the year, for which the statistic was calculated. The assets were grouped into five categories for the purposes of this study:

- Dwellings (residential buildings);
- Infrastructure, i.e. non-residential buildings and structures in ‘transportation and storage’ category (NACE sector H)⁶;
- Agricultural assets, i.e. non-residential buildings and structures, and machinery and equipment related to production in agriculture, forestry or fishery (NACE sector A), and cultivated biological resources;
- Industrial assets, i.e. non-residential buildings and structures, and machinery and equipment related to mining, manufacturing and utilities (NACE sectors B-E);
- Services assets, i.e. non-residential buildings and structures, and machinery and equipment related to other economic activity (NACE sectors F-U), and weapons systems, except assets under “infrastructure” category.
- Consumer durables, i.e. certain durable items bought by households for final consumption, such as furniture, electronics or personal vehicles.

Value of dwellings, infrastructure and consumer durables was calculated and inserted into the database as a relative value, in % of GDP. For the remaining three categories, their value was calculated relative to GDP generated by corresponding categories of production – agriculture (NACE sector A), industry (sectors B-E), and construction and services (sectors F-U).

As a result of preparing the fixed asset data, an update of the dataset from Paprotny et al. (2020) was generated and stored in the Zenodo repository. It contains standardized building structure and household contents value per m² of useful flood space of residential buildings for 31 countries for years 2000-2020.

Compared to V1: Consumer durables were added to list of fixed assets, thanks to new data by Paprotny et al. (2020).

3.3.9. Conversion of original damage values in monetary terms

Damage data in monetary terms need to be converted from their original, nominal values, to one currency and deflated to a single reference year. As with the GDP data (section 3.3.5), 2011 was chosen as the reference year and the currency is euro (EUR). A list of currencies was prepared for all countries and the entire period of the study. Its format is as follows:

Table 8. Currency database format

Column	Description
Code	NUTS0 two-letter country code
Name	Country/territory name
Currency	Currency name*

⁵ With an assumed service life of 11 years (Schmalwasser et al. 2011)

⁶ This category is generally intended to represent the value of roads, railways, airports, harbours and the like.

Code1	Three-letter currency code*
Code2	ISO 4217 numeric currency code
Start date	Date or year when currency first entered circulation
End date	Data or year when currency was withdrawn from circulation
Conversion	Conversion factor between new and old currency
Note	Other information relevant for correctly applying the information on currencies

Notes: * the currency name/code equals ISO 4217 currency name/code if the field 'Code2' is filled; otherwise the name/code is assigned solely for the purpose of disambiguation of different currencies in this database.

The data on currencies were mostly collected from ISO 4217 standard (ISO 2015) and Taylor (2004), amended from Internet sources. The conversion factors enable conversion from old to current currencies, and then to EUR where necessary, according to 2020 exchange rates reported in a separate table in the database (*Currency conversion*). Another table (*Deflator*) reports the values of deflators used to adjust nominal losses to real losses (2020 prices). The GDP deflator was generally used, as it allowed to make the loss adjustments consistent with GDP values. Only if the GDP was not available, alternative price indices were used, always “anchored” to the GDP deflator series. These series include indices of consumer prices, wholesale prices, retail prices or cost-of-living. The source of the data was usually the same as those for the GDP data. Some natural hazards databases, such as EM-DAT, report losses in US dollars, therefore exchange rates were obtained at *ad hoc* basis to convert those values to national currencies, usually by utilizing the same sources as for GDP or deflator series. It should be noted that the currency conversions and deflators omit six cases of hyperinflation: Germany 1922-23, Poland 1923, Greece 1942-45, Hungary 1945-46, Bosnia and Herzegovina 1992, and Yugoslavia (Serbia/Montenegro/Kosovo) 1992-93. Inclusion of those cases would cause large distortions to the data series. Hyperinflation periods and resulting currency changes were marked in the dataset. The dataset also includes deflator for five countries that do not exist anymore, but some regions or countries in the domain were part of in the past, namely Czechoslovakia, the Soviet Union and Yugoslavia.

An example calculation is shown below for the sake of illustration. It shows the conversion of an estimate of losses due to the 1934 flood in southern Poland:

- Losses in 1934 currency and prices: 74.6 million pre-war Zlotys;
- Pre-war Zloty was converted to “Cracow” Zloty in 1940 at par (1:1), again at par to post-war Zloty in 1945, then denominated to “heavy” Zloty (PLZ) at 100:3 in 1950, and finally to “new” Zloty (PLN) in 1995 at 10,000:1. Additionally, the exchange rate to euro (EUR) in 2020 was 4.443, hence

$$74600000 / \frac{1}{1} / \frac{1}{1} / \frac{100}{3} / \frac{10000}{1} / 4.443 = 50.37 \quad (9)$$

- Therefore, the uninflated value of losses equals 50.37 EUR. From the GDP deflator series we can extract the price index for 1934, which is approx. 0.0000242, where year 2020 equals 100. Therefore:

$$50.37 \times \frac{100}{0.0000242} = 208374117 \quad (10)$$

- Hence, the losses from the 1934 flood in 2020 prices and exchange rates can be estimated at 208.4 million EUR.

Compared to V1: Reference year from changed from 2011 to 2020. Some historical conversion rates were corrected, mainly for Central-European countries before 1950, including the example.

3.4. Land use and population distribution modelling

In this section the methodology of reconstructing temporal changes in land use and population distribution in Europe is described. In the simulation, computation of land use for a given year was done in turns for each land use class, as follows:

1. Special cases (3.4.15)
2. Sub-regional population redistribution (3.4.1);
3. Urban fabric and urban population redistribution (3.4.2);
4. Airports (3.4.6);
5. Reservoirs (3.4.13);
6. Rural population redistribution (3.4.3);
7. Industrial or commercial units (3.4.4);
8. Infrastructure (3.4.5);
9. Construction sites (3.4.7);
10. Green urban and sport areas (3.4.8);
11. Croplands (3.4.10);
12. Pastures (3.4.10);
13. Burnt areas (3.4.11);
14. Natural areas (3.4.12);
15. Soil sealing degree adjustment (3.4.16);

The procedure is carried out separately in each NUTS 3 regions, and then the results are merged to create maps of land cover/use and population. After the redistribution of land cover/use is done, the map of soil sealing is modified based on the transitions that have occurred.

3.4.1. Sub-regional population redistribution

Redistribution of population within NUTS regions was modelled to capture population changes observed in Europe in the past decades in three broad areas:

- Declining population in city “cores”, i.e. most centrally located and most densely populated parts of cities, especially large ones;
- Rapid growth of suburban zones around city cores;
- Declining population of rural areas.

The first two changes are largely driven by the change in number of persons per households. Even when the population of a city is stagnant, smaller families in each dwelling result in an increased demand for housing. Between 1870 and 2011, the number of urban households has increased 8-fold. Those extra dwellings had to be constructed mostly outside the urban centres, as existing houses are rarely replaced by bigger ones. Since the late 19th century many authors noted the functional relationship between population density and distance from the city centre (Berry et al. 1963, Anas et al. 1998, Papageorgiou 2014, Li et al. 2021). Clark (1951) has shown that over time, the sharp decline in population density with distance has become much less pronounced. This is largely caused by the aforementioned social change: in the existing households families become smaller, hence the population declines closer to the centre and the surplus population has to be accommodated in a larger distance from the centre, in less-developed areas. At the same time, migration from rural to urban areas has reduced population in the former and contributed further to the growth of suburbs.

The changes are clearly visible in available sub-regional data. The European Commission provides tabular data on population at the level of local administrative units, or LAUs, for 1961-2011 (Gløersen and Lüer 2013). After merging this dataset with administrative boundaries from EuroGeographics (see Appendix 3 for details on compiling the data), detailed population change patterns are revealed (Fig. 9). The correlation of this population change with several datasets was analysed. The potential explanatory variables included:

- Population density in different years from the LAU dataset and GIS-computed area;
- Population density in 2011 computed using kernel density on the 1 km GEOSTAT population grid (see section 3.2.3) and a 10-km radius, later called “agglomeration density”;
- Euclidean distances from urban centres: (a) Arbitrary centres of large agglomerations (more than 300,000 persons in 2018) and capital cities (United Nations 2018); (b) Centroids of population clusters calculated from the 1 km population grid (Eurostat 2021b, see section 3.2.3); (c) Centroids of high-density population clusters from the same source; (d) Centroids of cities included in the Urban Atlas 2018 (Eurostat 2021b);⁷ (e) Centroids of Corine Land Cover 2012 urban patches (see section 3.2.1).
- Total population of nearest city/population cluster/CLC urban patch⁸;
- Combined total population and inverse of Euclidean distances.

Further, data was analysed using original values and also normalized values (using country-level averages) and only for NUTS3 regions with at least 10 LAUs. The strongest correlation (0.68) was found between population growth relative to national average and population density in 2011. For all variables, strongly non-linear behaviour was found. As expected, for low population densities the correlation with population growth is strongly positive, while for high densities, corresponding to urban areas, the correlation is negative. Similarly, the growth is positively correlated with increasing distance from urban centers up to a certain distance, above which correlation turns negative.

⁷ For some non-EU countries not covered by this dataset, high-density clusters from (c) were added

⁸ Population was assigned to CLC patches from the disaggregated population grid (see section 3.2.1)

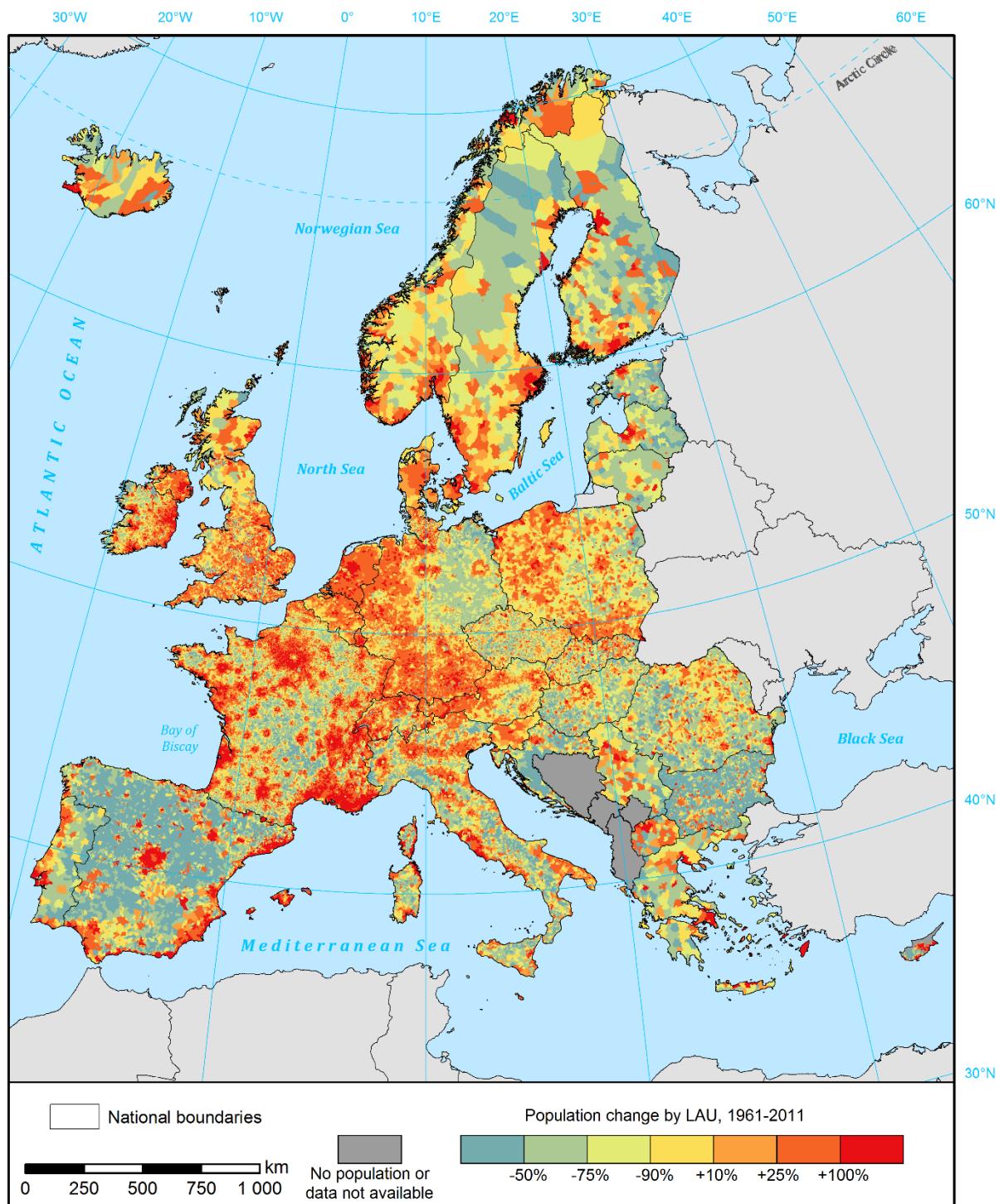


Figure 9. Population change by local administrative units, 1961-2011. See Appendix 3 for sources of data.

To capture the difference in growth patterns of high- and low-density LAUs, and further obtain the uncertainty of the model, two copulas were used (Fig. 10):

- Gaussian copula for LAUs with a population density below 2000 persons per km^2 that correlates population density from LAU data with population growth (Spearman's $r = 0.69$) and
- Frank copula for LAUs with a population density above 1500 persons per km^2 that correlates “agglomeration density” with population growth (Spearman's $r = -0.36$).

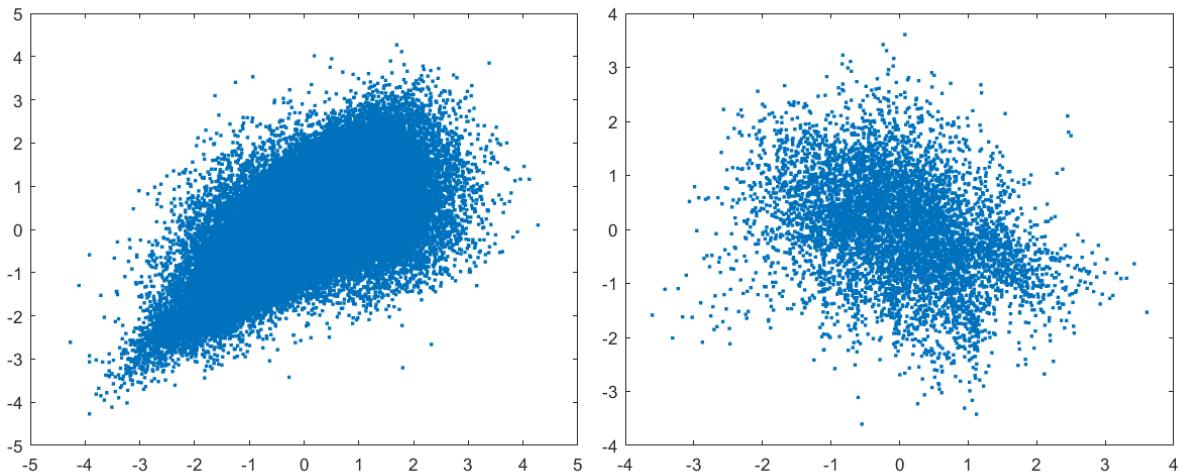


Figure 10. Empirical copulas of the dependency between population growth (1961-2011) relative to national growth, and population density (2011) in two different forms (left – local density, right – agglomeration density). Left copula for local population densities below 2000 persons per km² and right on for population densities above 1500 km²

The copulas were chosen based on a goodness-of-fit test (Genest et al. 2009). The Gaussian copula is applied for LAUs with population density of no more than 1500 km², and the Frank copula above that threshold. The value was selected as it gave the best results, and coincidentally it is also the threshold used to define high-density population clusters by Eurostat, and in turn to classify LAUs as urban.

LAUs used in the model are those from the Eurostat dataset, as they don't have complete coverage, the data have lower geometric accuracy, and size of LAUs varies substantially between countries. Therefore, a set of “virtual” LAUs was constructed. Every “virtual” LAU consists of an urban patch from Corine Land Cover 2012 and its nearest neighbourhood. Disaggregated population in 100 m resolution was then assigned to each “virtual” LAU to compute population density. For each timestep in the model, the copulas are sampled 10,000 times to obtain an estimate of annual population growth (geometric average). The population $P_{t,vLAU}$ of a “virtual” LAU in year t and 2011 is then:

$$P_{t,vLAU} = P_{2011,vLAU}(1 + A)^{2011-t} \quad (11)$$

where A is the annual growth rate (in %) from the copula model. To avoid unrealistic changes mainly for very low population density, population growth is capped: $-2.257\% < A < 1.6464\%$, which corresponds to a 10-fold decline or 25-fold increase between 1870 and 2011. The analysis is done separately for each NUTS region, hence parts of “virtual” LAUs in different regions are considered separate “virtual” LAUs. After the population of all “virtual” LAUs was computed for a given NUTS region, it is adjusted proportionally so that the combined population of “virtual” LAUs matches the population of NUTS regions in a given year as defined in the input database. The computation is done separately for urban and rural areas.

Compared to V1: This step of the modelling procedure is entirely new.

3.4.2. Urban population and urban areas (CLC 111-112)

After population is redistributed between “virtual” LAUs, urban population within each unit is redistributed further, in order to better model the process of suburbanization. The modelling approach is as follows:

1. In every urban fabric grid cell g in region (“virtual” LAU) r the population P in year t is modified relative to 2011 baseline to account for change in household size:

$$P_{t,r,g} = P_{2011,r,g} \frac{H_{t,r}}{H_{2011,r}} \quad (12)$$

where $H_{t,r}$ is the average number of persons per household, determined for each NUTS3 region (see section 3.3.5);

2. All grid cells in a region are ranked by distance from urban centres (explained further in the text), where the highest-ranked cells are the closest to any urban centre.
3. Surplus population S_t is calculated:

$$S_{t,r} = U_{2011,r} \frac{H_{t,r}}{H_{2011,r}} - U_{t,r} \quad (13)$$

where $U_{t,r}$ is the total urban population in the “virtual” LAU (see previous section);

4. The final step depends on the value of both S_t and t :
 - a. If $S_t > 0$ and $t < 2011$, it means that the urban area in year t was smaller relative to baseline, as the growth in number of households between t and 2011 was achieved through urban expansion. Urban grid cells are iteratively removed starting with the lowest-ranked, and their population is reduced by proportion $D_{r,g}$:

$$P_{t,r,g} = P_{t,r,g} D_{r,g} \quad (14)$$

until the urban population in the region matches the desired value of $U_{t,r}$. Proportion $D_{r,g}$ is based on the logarithm of distance from urban centres d (in hectometres):

$$D_{r,g} = \left(1 - \frac{\ln(d_{r,g})}{\ln(\arg \max_{d_{r,g} \in d_r} \{d_{r,g}\})} \right) \quad (15)$$

If at some point there is more population in grid cell(s) than remaining surplus, the population is reduced by the available amount, split proportionally to grid cell population if there are more cells with the same rank.

- b. If $S_t > 0$ and $t > 2011$, it means that the number of households in the urban area declined between 2011 and t as dwellings became vacated, hence the urban area remained unchanged. The population in all urban grid cells is reduced proportionally until $\sum P_{t,r,g} = U_{t,r}$.
- c. If $S_t < 0$ and $t > 2011$, it means that the urban area in time year t was larger relative to baseline, as the growth in number of households between 2011 and t was achieved through urban expansion. Cells where the expansion most likely took place are identified using a land-use transition model described in section 3.4.10, starting with cells with the highest probability of transition. If more cells were given the same likelihood of transition to urban fabric than necessary to assign the additional population, the cells within that group were ranked according to distance from the urban centre. The population in the highest-ranked cells is set to the maximum population per grid cell in the “virtual” LAU, reduced by proportion $D_{r,g}$:

$$P_{t,r,g} = \arg \max_{P_{t,r,g} \in P_{t,r}} \{P_{t,r,g}\} D_{r,g} \quad (16)$$

If at some point the remaining surplus is lower than the aforementioned value, or there are no more empty grid cells available, the surplus is distributed equally between all cells that were modified until this step. If there are no available grid cells at all in the “virtual” LAU, the population of all urban grid cells is increased proportionally until $\sum P_{t,r,g} = U_{t,r}$.

- d. If $S_t < 0$ and $t < 2011$, it means that the number of households in the urban area declined between t and 2011 as dwellings became vacated, hence the urban area remained unchanged. The population in all urban grid cells is increased proportionally until $\sum P_{t,r,g} = U_{t,r}$.

The important aspect influencing the result of this process is the “distance from urban centre”. Urban networks have several levels of hierarchy, with large agglomerations influencing population distribution far outside their borders. Euclidean distances from urban centres from five datasets described in section 3.4.1 were tested. Based on a calibration process, which is explained in Appendix 4, four of those datasets were selected for combined distance from urban centres. Each of them was given a different weight:

- Arbitrary centres of large agglomerations and capital cities, with a weight of 1.0;
- Centroids of high-density population clusters, with a weight of 1.5;
- Centroids of cities included in the Urban Atlas 2018, with a weight of 2.0;
- Centroids of Corine Land Cover 2012 urban patches, with a weight of 0.5.

Transition from/to urban fabric due to decrease or increase in population was calculated based on total observed changes in the study area according to CLC inventories. Between 2000 and 2012, urban fabric expanded by almost 1.88 million ha, while between 2012 and 2018 only by 98,676 ha, according to the CLC data. Consequently, urban fabric is removed in timesteps before baseline year 2011 only if the population in a grid cell was reduced to less than 9 persons. For timesteps after 2011, only an increase of population to more than 81 per 100 m grid cell resulted in transition to an urban fabric class.

Compared to V1: the model was recalibrated given the new sub-regional modelling step, using also some new “distance to urban centre” datasets. Urban population is now reduced gradually, rather removed altogether on the edges. Modelling urban expansion is now done using a BN for land-use transitions.

3.4.3. Rural population redistribution

Rural population is recalculated separately for each “virtual” LAU, by adjusting the grid cell baseline population proportionally to its value, so that it equals the expected population in that area (see section 3.4.1)⁹. For years before the baseline, areas from which urban fabric was removed are still considered urban for the purpose of this calculation. For years after the baseline, rural population in areas that transitioned to urban fabric in areas is no longer considered rural, hence the adjustment is made only to remaining rural cells in the “virtual” LAU.

As the population is always recorded as integers, a discrepancy might arise from adjusting the population in grid cells and then rounding it. Therefore, the adjusted population in grid cells is first rounded using “floor” function, and then the grid cells with the highest remainders from dividing the unrounded values by 1 are rounded using “ceiling” function. The number of highest remainders is determined by the difference between the expected population in the “virtual” LAU and the total population of cells adjusted and rounded using “floor” function.

Compared to V1: due to the new sub-regional modelling step, the rural population is adjusted proportionally to the grid cell value, rather than redistributed similarly to urban population.

⁹ For timesteps after the baseline year, population living in areas taken up by an airport or reservoir is removed first.

3.4.4. Industrial or commercial units (CLC 121)

The area covered by large industrial/commercial facilities was assumed to change proportionately to GDP generated by industry and services in constant prices (see section 3.3.7). Industrial grid cells located furthest from the centroids of industrial land use patches are removed first when going back in time. For timesteps after the baseline year, industrial grid cells closest to the centroids are added first. Industrial land use is only allowed to spread into uninhabited cells of some CLC classes: construction sites (133), agricultural (211–244) and certain natural (311–324 and 333).

However, growth in GDP from industry and services is only partially driven by expansion of facilities, as the productivity of capital and labour tends to increase. Indeed, CLC 121 class has grown between 2000 and 2018 (based on CLC 2012 and CLC-Changes) by 16%, but GDP from industry and services increased by 32%. Therefore, the change in GDP from industry/services is scaled by elasticity of 0.45, so that modelled changes between 2000–2018 in the study area have the same magnitude as observed in the CLC inventory. The industrial area A in region r and year t is as follows:

$$A_{r,t} = A_{r,2011} \left(\frac{G_{r,t}}{G_{r,2011}} \right)^{\varepsilon} \quad (17)$$

where ε is the elasticity and G_r is the GDP from industry/services according to the NUTS3 database.

Compared to V1: total GVA from industry is used as proxy, rather than a per-capita value. Elasticity was added to account for the sensitivity of land use change to GVA change.

3.4.5. Road and rail networks and associated land (CLC 122)

The area covered by roads and railways before year 2000 was assumed to change proportionately to the length of motorways and railways (see section 3.3.4). Infrastructure grid cells located furthest from the urban centres are removed first when going back in time until the total area per region matches the value in the NUTS3 database. Conversely, grid cells closest to the urban centres are filled with infrastructure for timesteps after the baseline year. Infrastructure is allowed to spread only in certain CLC classes, as for industry. However, construction sites were prioritized over other CLC classes; all ‘construction’ grid cells have to be used up before other CLC classes can be considered. This is done because almost half of the area of construction sites in the CLC inventory that transitioned to class other urban fabric or industrial sites (considered in previous steps) transitioned to infrastructure by the time of the next 6-year inventory.

Compared to V1: prioritization of construction sites for timesteps after the baseline year was added.

3.4.6. Airports (CLC 124)

Airports first appeared in early 20th century. Due to the relatively small number of those objects in Europe (1,598) and mostly well-described history, a given airport was entirely removed from the land use dataset using information on the year of construction. This approach assumes that the area of the airport hasn't changed since its foundation; the assumption seems valid for most airports, however. Airports were identified mostly by intersection of Corine Land Cover 2012 data with OurAirports (2021) open database, while the year of construction was gathered from various Internet resources (mostly Wikipedia). In some cases, the construction year was not available, therefore it had to be estimated based on available information, such as circumstances of their construction or runway parameters¹⁰.

¹⁰ For instance, a substantial part of European airports was built in the run-up to World War II (1930s) and during the military build-up of the early Cold War (1950s). Interwar and World War II airports typically have two, parallel,

Nine new airports appeared in CLC 2018; those airports are added to the baseline map according to the year of construction. If there is any population in the area where the airport would be built, it is removed. Fig. 11 shows the number of airports identified, by year of construction.

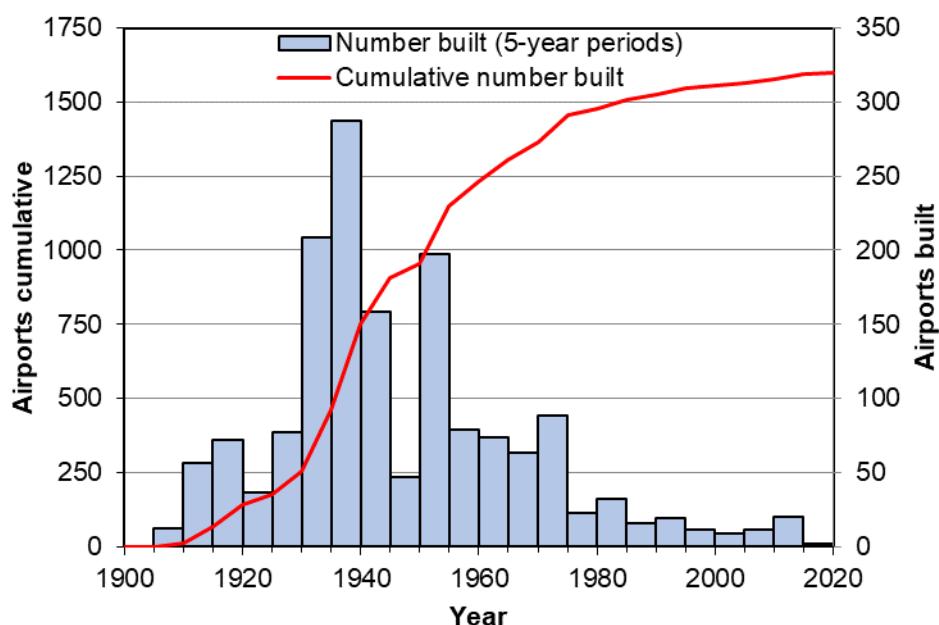


Figure 11. Number of airports built in the study area between 1900 and 2020 (5-year periods ending with the year indicated).

Compared to V1: list of airports was revised based on the revised CLC 2012 layer and updated with new OurAirports data and other Internet resources. Airports for new countries and new airports from CLC 2018 were added.

3.4.7. Construction sites (CLC 133)

Construction sites are by definition a temporary land use, typically for only a few years. The CLC inventory shows that 76–81% of construction sites transition to another land use during the 6-year periods between CLC datasets. Therefore, for years 2005–2011, their area was assumed constant, while for years 1870–2004 all construction sites were removed from the dataset. After 2011, construction sites were allowed to transition into urban fabric, industrial sites, roads, railways and airports (CLC 111–122 and 124), as described in sections 3.4.2–3.4.5, but otherwise kept unchanged.

Compared to V1: certain artificial surfaces can now expand into construction sites after baseline year (previously construction sites were kept constant).

3.4.8. Green urban areas, sport and leisure facilities (CLC 141 and 142)

Green urban areas, sport and leisure facilities are closely related to other artificial surfaces. Indeed, almost two-thirds of those CLC patches border either urban fabric, industrial sites, road/railway sites, or airports in the CLC 2012 inventory. Therefore, those patches of land cover in CLC 141–142 class which bordered CLC classes 111–122 and 124 in the baseline map are removed if in a given timestep that was no longer true.

1–1.5 km long runways, one paved and one grassy; Cold War-era military airports have usually one or more, long (about 2.5 km), paved runways, often intersecting with each other.

Compared to V1: new modelling process added, as described (previously assumed constant).

3.4.9. Port areas, mineral extraction sites, dump sites (CLC 123, 131 and 132)

Ports, mineral extraction and dump sites constitute between themselves less than 0.2% of total land area in Europe, and it would be very difficult to collect the year of construction for each of almost 15,000 objects. Therefore, their area was kept constant at every time step. Similarly, those classes do not interact with other land use classes. However, in rare cases, they might be replaced in certain timesteps by water bodies (CLC 512) in case of reservoir or polder construction (see sections 3.4.13 and 3.4.15).

3.4.10. Agriculture (CLC 211-244)

Changes in agricultural area are computed primarily using a Bayesian Network (BN) model quantified with observed changes between CLC inventories from 2000 to 2018. In the three CLC-Changes datasets available for that time period, almost 1.2 million transitions involving patches of land larger than 5 ha were observed. CLC-Changes and CLC 2012 inventory were sampled to obtain 513,915 cases of transition and equal number of land-use being stable between 2000 and 2018. For each location, information from different raster datasets (on terrain, agricultural suitability, population density and urbanization) were extracted as predictors of land-use changes. The sampling procedure and a list of all tested predictors is described in Appendix 5.

As the land use information is categorical, a discrete BN was used. The BN model was constructed iteratively, starting with a simple three-node network, where the “old” land-use class is the parent of “new” land-use class, and a single predictor variable is the parent of both land-use nodes. The model with a different predictor was validated against a smaller dataset containing samples of transitions and non-transitions (see Appendix 5). Iteratively, the best variables were selected, and also the optimal numbers of bins into which the continuous variables were discretized were also selected. Three predictors were chosen, all of which are parents of the two land-use nodes (“new” or “old” land-use is also a predictor, depending which is the variable of interest for a particular application). They are (Appendix 6, Figs. A4–A6):

- Population density per “virtual” LAU (see section 3.4.1) – 9 bins;
- Agricultural suitability indices for 1971-2000 calculated by FAO (2021b) in the Global Agro-Ecological Zoning version 4 (GAEZ) database:
 - Output density (potential production divided by total grid cell area) for wheat under rainfed conditions and high input level – 5 bins;
 - Agro-climatic potential yield for grass with an available water content of 200 mm/m (under irrigation conditions) and high input level – 10 bins.

As the BN is quantified with a conditional probability table (CPT), this configuration results in the CPT having 11,250 cells. Therefore, no more variables were added to avoid too few data points quantifying cells of the CPT.

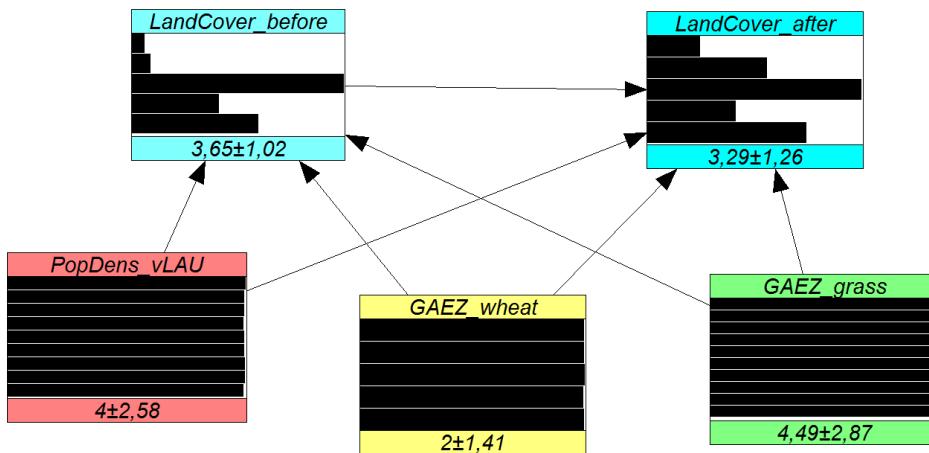


Figure 12. Bayesian Network for land-use transitions, visualized using UniNet software (<https://lighttwist-software.com/uninet/>).

The BN is used to generate probabilities of land-use transitions in nine cases, as follows:

- from non-urban to urban after the baseline year;
- from non-cropland to cropland after the baseline year;
- from non-pasture to pasture after the baseline year;
- from cropland to non-cropland after the baseline year;
- from pasture to non-pasture after the baseline year;
- from non-cropland to cropland before the baseline year;
- from non-pasture to pasture before the baseline year;
- from cropland to non-cropland before the baseline year;
- from pasture to non-pasture before the baseline year.

The case of non-urban to urban transition is described in section 3.4.2. All others are used in the same way, i.e. they determine in which areas add or remove agricultural areas so that the total area of croplands and pastures in the land cover map for a given timestep matches the values obtained from historical statistics (see section 3.3.6). This is done iteratively starting with patches of land with the highest probability of transition between given classes (e.g. non-pasture to pasture). Transition of artificial surfaces to agricultural land-use is not allowed, and similarly transition from cropland to pasture is not allowed (only the other way round)¹¹. Non-utilizable land cover types (CLC 331-332, 334-335, 421-523) are excluded entirely from interactions with croplands.

In the iterative land-use redistribution process, the number of grid cells with equal probability of transition might exceed the number of cells that need to be converted to match the total area in the historical statistics. This often happens as the predictors (GAEZ dataset and population density by “virtual” LAU) have relatively coarse resolution together with a small number of bins in which the data

¹¹ When modelling timesteps before 2011, the chronology is reversed, therefore when urban fabric or roads/railways sites are partially removed in cases when it occupied less land in the past, an empty space is created, which croplands can occupy. In the actual chronology, it is a cropland to non-cropland transition. The probability of transition that is used to determine which cells to fill with croplands is the probability that a non-cropland cell was a cropland before. Same applies to pastures, but they cannot occupy cells already assigned to cropland.

is divided. In such case, the cells are ranked according to slope of the terrain (from lowest to highest). This variable was calculated from EU-DEM elevation dataset at 100 m resolution (Eurostat 2021b, see Appendix 6, Fig. A7). The appropriate number of highest-ranked cells is added/removed so that the total area of croplands or pastures matches the total area in the historical statistics.

The crop suitability indices do not change over time, thus disregarding any variability of climate conditions, though Klein Goldewijk et al. (2011) considered this to be a valid assumption despite a much longer timespan in the HYDE database. That study also used agricultural suitability and slope to predict where croplands and pastures needed to be added/removed in order to reconstruct total area of those land uses in different countries in historical periods.

Compared to V1: deterministic functions used to rank cropland and pasture cells in the previous version were replaced by a discrete Bayesian Network, though the predictors (agricultural suitability, slope) are similar.

3.4.11. Burnt areas (CLC 334)

Areas where vegetation has burned down (typically forests) are by definition a temporary land use. It is very short-lived: almost none of the land cover patches in this class in 2012 were present in either 2006 or 2018 CLC inventories. For years 2007–2017, burnt area was assumed constant, while for years 1870–2006 and 2018–2020 all burnt areas were removed from the dataset. As almost all burnt areas are formerly or subsequently CLC classes 311–324, this modelling step is done after redistributing agricultural areas.

Compared to V1: timestep during which burnt areas are assumed constant was slightly adjusted based on new CLC data.

3.4.12. Vegetated natural areas (CLC 311–324, 333 and 411–422)

The remaining land use after subtracting artificial, agricultural and burnt areas are natural areas, which would have covered the entire continent without human activity. Therefore, if as a result of the land use modelling some land becomes unoccupied (e.g. not used for housing or agriculture in a given time step), it is assumed that this land was covered by the same natural land cover that is typical in its nearest neighbourhood. “Typical” natural vegetated land cover (classes CLC 311–324 and 411–422) is defined as the most frequently occurring one within the “virtual” LAU (see section 3.4.1) a given cell belongs to. The calculation is done separately for forests (CLC311–313) and other natural land (CLC 321–324 and 411–422), and the more frequent of the two is used. If there is no natural cover in the “virtual” LAU, the dominant vegetated land cover of the applicable NUTS3 region is used. If no vegetated land cover was located in the vicinity, the unoccupied land was assumed to be covered by transitional woodland-shrub (CLC 324), as it is the most common non-forest natural land.

After the first allocation, the total area of forests is compared with the historical data in the NUTS3 database. If there is too much forest area in a given NUTS3 region, the land that was allocated to forest in this step is iteratively converted into the most frequent non-forest class, starting with the most-densely populated “virtual” LAU. Conversely, if there is not enough forest land, cells that were allocated to non-forest vegetation in this step are iteratively converted into the most frequent forest class, starting with the least-densely populated “virtual” LAU.

Compared to V1: CLC classes that are considered when filling empty space were adjusted, and the “nearest neighbourhood” was redefined. Also, forest area statistics are now considered during the allocation to achieve a better match with historical data.

3.4.13. Areas covered by water, incl. intertidal flats (CLC 423 and 511–523)

Areas covered by water are assumed to be constant over time, thus not allowing for coastline and river course changes etc., except for large reservoirs. A given reservoir was entirely removed from the land use dataset using information on the year of construction. 1121 reservoirs and their construction year were identified by intersecting Corine Land Cover 2012 and 2018 with the Global Reservoir and Dam (GRanD) Database v1.3 (Lehner et al. 2011, Global Dam Watch 2021). Reservoirs which are former or still largely natural lakes are excluded. Multiple reservoirs drawn together in CLC were considered together. Three dams built after 2011 and identified in CLC 2018 are added to the baseline map according to the year of construction. If there is any population in the area where the reservoir would be built, it is removed. Fig. 13 shows the number of reservoirs identified, by year of construction.

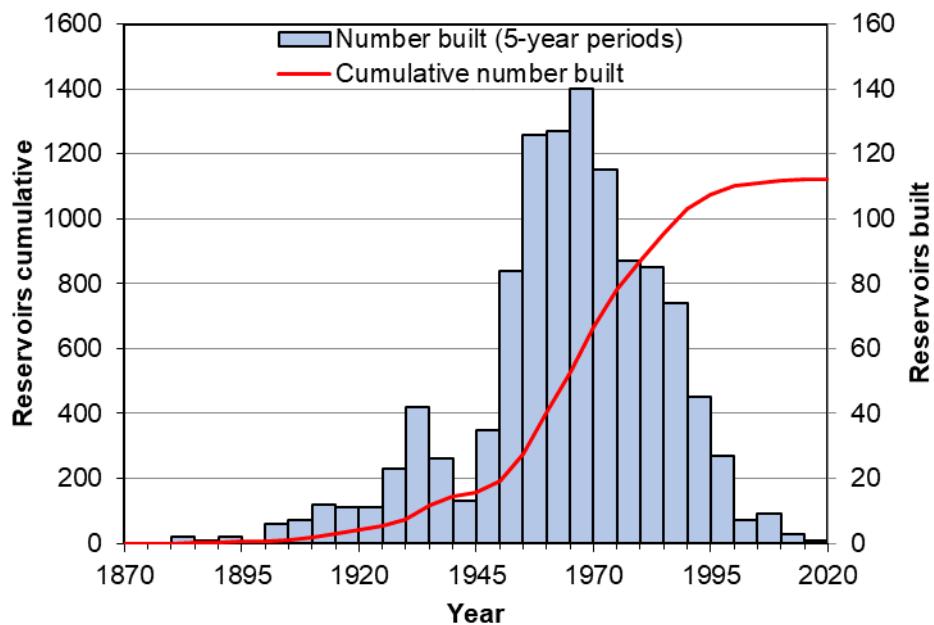


Figure 13. Number of large reservoirs built in the study area between 1870 and 2020 (5-year periods ending with the year indicated). Based on data from Global Dam Watch (2021).

Compared to V1: list of reservoirs was recompiled from scratch based on revised CLC 2012 inventory and new version of GRanD database (v1.3 from 2019 instead of v1.1 from 2011). Reservoirs for new countries and new reservoirs from CLC 2018 were added.

3.4.14. Other natural areas (CLC 331–332 and 335)

Beaches, dunes, sands (CLC 331), bare rocks (CLC 332), glaciers and perpetual snow (CLC 335) are not allowed to interact with any other classes, therefore they are assumed constant throughout.

Compared to V1: the CLC classes in this subsection are no longer considered when gap-filling empty land.

3.4.15. Special cases

A special case that has a rather large influence on exposure distribution in the Netherlands is considered, namely the *Zuiderzeewerken*. It was a large land-reclamation and flood-protection project, which resulted in the construction of large dikes and polders in the Zuiderzee between the 1920s and 1970s (Fig. 14). Zuiderzee was closed in 1932 by a large dike, turning into a lake¹². Cities, infrastructure

¹² The lake was further split by a dike in 1975, with the two parts being named IJsselmeer and Markermeer.

and farming land were created in the new spaces. Before 1942, the Flevoland province (created formally in 1986) consisted only of the small island town of Urk and the uninhabited island of Schokland. Therefore, the artificially-created land is removed from the land-use dataset and turned into inland water (CLC 512) for years before the year of completion. Population and “virtual” LAUs are also removed and not considered in the population and land-use redistribution for those years.

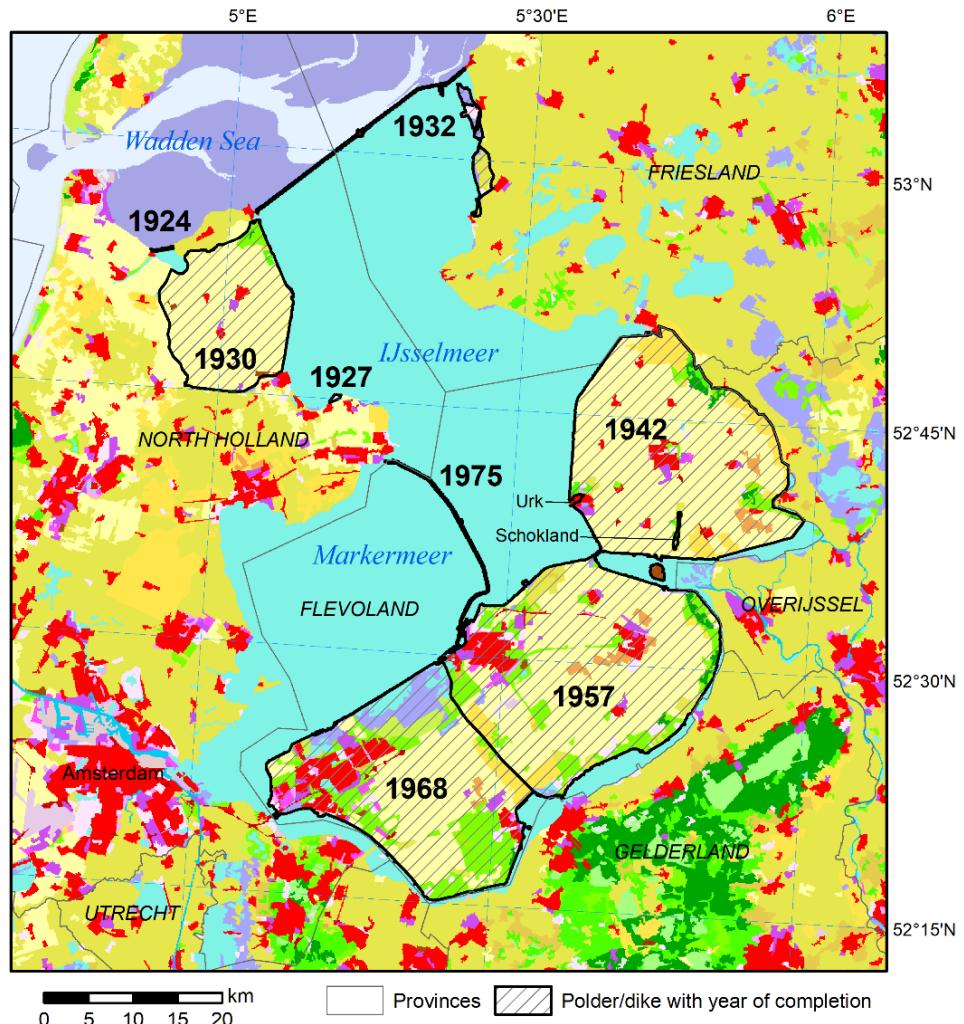


Figure 14. „Zuiderzeewerken“ land reclamations by year of construction, overlaying Corine Land Cover 2012 map (see section 3.2.2 for explanation of symbols).

Compared to V1: this component was added in V2.

3.4.16. Soil sealing

Changes in soil sealing are entirely based on land-use transitions. Soil sealing in the baseline map is increased to the average value for a given CLC class (Table 9) if non-artificial land transitions to artificial, unless it is already higher than that value. In the timesteps before the current one, wherever land that is currently artificial had been agriculture or natural, the degree of soil sealing is reduced to 1%. Similarly, it is reduced to 0% in cases when agricultural or artificial land was previously natural land in a given timestep.

Table 9. Average soil sealing in the Imperviousness 2012 dataset per selected types of Corine Land Cover classes.

CLC classes	Average soil sealing in 2012
Urban fabric (111–112)	28%
Industrial or commercial units (121)	45%
Road and rail networks (122)	29%
Airports (124)	20%
Agricultural areas (211–244)	1%
Natural land (311–523)	0%

Compared to V1: this component was added in V2.

3.5. Disaggregation of economic data

Disaggregation of economic data generates exposure estimates of GDP and wealth per grid cell, consistent with the population and land use data. The methodology presented here follows dasymetric mapping methods, particularly considering European studies such as EU's ESPON 2013 Programme (Milego and Ramon 2011), HARCI-EU dataset (Batista e Silva et al. 2019) or global studies such as Murakami and Yamagata (2019). Whereas many existing studies use only a single predictor, mainly population (Kummu et al. 2018) or night-time lights (Eberenz et al. 2020), here we combine population, land-use and soil sealing information.

Table 9 provides a summary of the assumptions behind the disaggregation. Regional GDP (see section 3.3.7) is split partially proportionally to population and partially according to land-use (with soil sealing where appropriate). This is done because all economic activities require both capital and labour input; capital is connected to land-use, while labour is part of the general population. Therefore, using the observation that the labour share in advanced countries is about 60% and has been relatively stable over time (ILO and OECD 2015)¹³, GDP is disaggregated 60% using population and 40% using land use.

Fixed assets in absolute terms are computed by multiplying regional GDP, or a sector thereof, by the appropriate wealth-to-GDP ratio for each sector, as defined in the database (see section 3.3.8). Fixed assets are distributed either according to population (housing and consumer durables) or land use proportionally to degree of soil sealing (all other sectors). There are also specific assumptions on disaggregation of GDP and wealth for specific sectors.

Agriculture and forestry. Additional assumption had to be made for the agricultural sector, which is the most dispersed, as almost three-quarters of the study area covered by agricultural land use or forests. At the same time, farmland and pastures are more productive and contain more assets than forests, especially since trees do not count as fixed assets. However, breakdown of GDP by agriculture and forestry is not available at regional level, and very limited historical data exist with such detail on national level. Very few countries show amount of fixed assets in the forestry sector. Hence, agricultural GDP and wealth at regional level were broken down to forestry (incl. logging) and remaining agriculture (incl. fishing and aquaculture) using an index of forest economy efficiency. This

¹³ Employee's compensation constitutes approximately half of GDP in European countries (Eurostat 2021a), however after accounting for self-employment the share is closer to 60%. Valuation also influences the estimates of labour share: at market prices it was about 55–65% in advanced countries in recent decades, but 65–70% at factor cost (ILO and OECD 2015). Market price is the currently used measure of GDP, though historical data are often at factor cost.

index was compiled by computing GDP from agriculture (without forestry) at national level¹⁴ per ha of agricultural land from CLC and GDP from forestry per ha of forest land. Those values were computed for the year 2000 for all countries and presented as efficiency of forest economy relative to other agriculture in %. This ratio was used to compute the relative share of forestry in regional GDP in any given year based on land cover/use modelled for that year:

$$G_{f,r,t} = G_{af,r,t} \frac{A_{f,r,t} E_{f,c}}{(A_{a,r,t} + A_{f,r,t} E_{f,c})} \quad (18)$$

where G is GDP, A is area covered by land cover/use in a particular sector, E_c is the efficiency index for country c . The forest sector is denoted by f , agricultural sector (without forestry) by a , NUTS3 region by r and timestep by t . Agricultural GDP without forestry is therefore:

$$G_{a,r,t} = G_{af,r,t} - G_{f,r,t} \quad (19)$$

60% of GDP is distributed proportionally to the population living in agricultural areas (CLC classes 211–244) or forests (CLC 311–313), while the other 40% equally among those land-use classes, without considering soil sealing. Wealth-to-GDP ratio for agriculture is used both forestry and other agriculture. The resulting wealth estimates are distributed proportionally to the degree of soil sealing in either agricultural areas or forests.

Industry and mining. As agricultural data, industrial GDP and wealth data require further disaggregation. Mining and quarrying are split from the remaining industrial activities (manufacturing and utilities) using a mining economy efficiency index, calculated as the forest index. In the same way, it uses the proportion of mining areas (CLC 131) relative to industrial areas (CLC 121) in each NUTS3 region and timestep to disaggregate the two sectors. Eq. 17 and 18 are applicable here, just with the different sectors.

60% of GDP is then distributed proportionally to the total population, while the other 40% is distributed to the appropriate land use classes. However, the degree of soil sealing is only considered for industrial sites; the distribution of GDP is proportional to the percentage of the soil sealed. Wealth-to-GDP ratio for industry is used both mining and other industry. The resulting wealth estimates are distributed proportionally to the degree of soil sealing in either agricultural areas or forests.

Services. This sector is less clearly attached to land-use. Transportation and storage activities can be attributed to infrastructure (CLC 122–124), while construction at least partially takes place at large construction sites (CLC 133). Other services, such as trade, accommodation, communication, real estate, health care, education or administration is related mainly with populated places, urban areas in particular. 60% of GDP from services is distributed proportionally to the total population, while the other 40% is distributed to several land use classes, proportionally to the degree of soil sealing: urban fabric (CLC 111–112), “industrial, commercial and transport units” (CLC 121–124), construction sites (CLC 133), and “artificial, non-agricultural vegetated areas” (CLC 141–142).

Fixed assets in services exclude the value of infrastructure, hence only CLC classes 111–121, 133 and 141–142 are used to disaggregate this category, again proportionally to soil sealing.

Infrastructure. Infrastructure is only shown separately among fixed assets, and follows a slightly different approach. The value is distributed first evenly among urban fabric (CLC 111–112) and “industrial, commercial and transport units” (CLC 121–124), but soil sealing is only used for classes that

¹⁴ NACE sectors A01 and A03. Detailed GDP breakdown by 64 sectors at national level only is available from Eurostat (2021a) for majority of countries in the study area.

represent purely infrastructure: road/railways sites, ports and airports (CLC 122–124). For the other classes consisting of urban fabric and industrial/commercial sites (CLC 111–121) the value of infrastructure is distributed proportionally to the percentage of land in a grid cell covered by roads and streets from the European Settlement Map 2012 (see section 3.2.3).

Residential assets. Residential assets consist of housing (residential buildings) and consumer durables (household contents) and are distributed proportionally to the population, regardless of land cover/use.

Table 10. Disaggregation of economic variables by population and land use classes (CLC = Corine Land Cover, see Fig. 3 and Table 1).

Variable	Category	Population	Land use	Soil sealing
GDP	Agriculture excl. forestry	Population in CLC211-244 (60%)	CLC211-244 (40%)	-
	Forestry	Population in CLC311-313 (60%)	CLC311-313 (40%)	-
	Industry excl. mining	Total population (60%)	CLC121 (40%)	yes
	Mining	Total population (60%)	CLC131 (40%)	-
	Services	Total population (60%)	CLC111-124/133/141/142 (40%)	yes
Wealth	Housing	Total population	-	-
	Consumer durables	Total population	-	-
	Agriculture excl. forestry	-	CLC211-244	yes
	Forestry	-	CLC311-313	yes
	Industry excl. mining	-	CLC121	yes
	Mining	-	CLC131	yes
	Services	-	CLC111-121/133/141/142	yes
	Infrastructure	-	CLC111-124	yes*

Note: * for CLC classes 111–121 using fraction covered by streets and roads only rather than the total degree of soil sealing.

A caveat of the approach of distributing regional GDP according to population at high resolution is that labour often resides outside the NUTS3 region where work is carried out. Therefore, too much GDP is assigned to the population of, in particular, capital cities or countries such as Luxembourg and Liechtenstein, and too little to the surrounding areas.

Compared to V1: GDP is now distributed by population and land-use in 60:40 proportion rather than 50:50. Mining is now distributed separately from industry, and soil sealing degree is introduced as an additional disaggregation factor. Lists of land-use classes per type of asset/activity were adapted.

4. HANZE-Events for floods: concepts and contents

HANZE-Events includes information on past damaging floods that occurred in the study area between 1870 and 2020. Records of flood events were obtained from a large variety of sources, including international and national databases, scientific publications and news reports, which are indicated per event in the HANZE-Events dataset. For detailed description of the inclusion criteria and definitions of

individual fields, and overview of the results, the user is referred to the paper in *Earth System Science Data*. An overview of the type of data collected is shown in Table 11.

Table 11. Information contained in HANZE-Events database.

Variable	Short description	Mandatory?	Field type	Permitted values
ID	Unique event identifier	yes	integer	1...9999
Country code	Two-letter country code	yes	string	Country codes
Year	Year of the event	yes	integer	1870...2020
Country name	Country name	yes	string	Country names
Start date	Daily start date	yes	date	1.1.1870...31.12.2020
End date	Daily end date	yes	date	1.1.1870...31.01.2021
Type	Detailed type of event	yes	string	River, Flash, Coastal, River/Coastal
Flood source	Name(s) of rivers, lakes or seas which caused the flooding	if available	string	Free text, with names delimited by semicolon
Regions affected (v2010)	Regions where human or economic losses were reported, at NUTS3 level (version 2010)	yes	string	Region codes
Regions affected (v2021)	As above, but at using NUTS version 2021	yes	string	Region codes
Area inundated	Area inundated (km ²)	at least 1 field on impact magnitude has to be recorded	integer	1...inf
Fatalities	Persons killed, incl. persons missing and presumed dead		integer	0...inf
Persons affected	Number of people evacuated or whose houses were flooded		integer	1...inf
Losses (nominal value)	Direct losses to assets in monetary terms in the currency and prices of the year and location of the flood event		integer	1...inf
Losses (original currency)	Three-letter code of currency in which nominal losses are recorded		string	Currency codes (see section 3.3.9)

Losses (2020 euro)	Direct losses to assets in monetary terms converted to euro in 2020 prices using the gross domestic product deflator	if nominal losses are recorded	integer	1...inf
Cause	The meteorological causes of the event	if available	string	Free text
Notes	Other relevant information or notes on issues with the data	if applicable	string	Free text
Data sources	List of publications and databases from which the information was obtained	yes	string	Free text, with citations delimited by semicolon
Changes	Indicates changes for HANZE v1	if applicable	string	Corrected, Updated, New

References

- Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P., and Feyen, L. (2014) Advances in pan-European flood hazard mapping. *Hydrological Processes* 28, 4067–4077. doi:10.1002/hyp.9947
- Anas, A., Arnott, R. and Small, K. A. (1998) Urban Spatial Structure. *Journal of Economic Literature*, 36(3), 1426–1464.
- Batista e Silva, F., Gallego, J., and Lavalle, C. (2013) A high-resolution population grid map for Europe. *Journal of Maps*, 9(1), 16–28. doi:10.1080/17445647.2013.764830
- Batista e Silva, F., Forzieri, G., Marin Herrera, M. A., Bianchi, A., Lavalle, C. and Feyen, L. (2019) HARCI-EU, a harmonized gridded dataset of critical infrastructures in Europe for large-scale risk assessments. *Scientific Data*, 6, 126. doi:10.1038/s41597-019-0135-1
- Berry, B. L. J., Simmons, J. W., and Tennant, R. J. (1963) Urban Population Densities: Structure and Change. *Geographical Review*, 53(3), 389–405. doi:10.2307/212588
- Bolt, J., Inklaar, R., de Jong, H., and van Zanden, J. L. (2018) Rebasing ‘Maddison’: new income comparisons and the shape of long-run economic development. GGDC Research Memorandum GD-174, Groningen Growth and Development Centre, University of Groningen, <https://pure.rug.nl/ws/portalfiles/portal/53088705/gd174.pdf>
- Clark, C. (1951) Urban Population Densities. *Journal of the Royal Statistical Society: Series A*, 114(4), 490–496.
- Clark, C. (1967) *Population Growth and Land Use*. Macmillan, London.
- Copernicus (2017) CLC 2012 final validation report. <https://land.copernicus.eu/user-corner/technical-library/clc-2012-validation-report-1>
- Copernicus (2019) HRL Imperviousness Degree 2015 validation report. <https://land.copernicus.eu/user-corner/technical-library/hrl-imperviousness-2015-validation-report>
- Copernicus (2021) Pan-European, <http://land.copernicus.eu/pan-european/>. Last accessed 23 July 2021.

- Eberenz, S., Stocker, D., Röösli, T. and Bresch, D. N. (2020) Asset exposure data for global physical risk assessment. *Earth System Science Data*, 12, 817–833. doi:10.5194/essd-12-817-2020
- Eicher, C. L. and Brewer, C. A. (2001) Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, 28(2), 125–138.
- European Environment Agency (2011) Legend, <http://www.eea.europa.eu/data-and-maps/figures/corine-land-cover-2006-by-country/legend>. Last accessed 8 December 2016.
- European Union (2011) Commission Regulation (EU) No 31/2011 of 17 January 2011 amending annexes to Regulation (EC) No 1059/2003 of the European Parliament and of the Council on the establishment of a common classification of territorial units for statistics (NUTS), OJ L 13, 18.1.2011, 3–54.
- European Union (2013) European system of accounts ESA 2010. Publications Office of the European Union, Luxembourg, doi:10.2785/16644, 688 pp.
- Eurostat (2021a) Database, <http://ec.europa.eu/eurostat/data/database>. Last accessed 22 September 2021.
- Eurostat (2021b) GISCO - the Geographic Information System of the COmmission, <https://ec.europa.eu/eurostat/web/gisco/overview>. Last accessed 21 October 2021.
- FAO (2021a) FAOSTAT, <http://www.fao.org/faostat/en/#data/RL>. Last accessed 22 September 2021.
- FAO (2021b) Global Agro-Ecological Zoning version 4 (GAEZ v4), <https://gaez-data-portal-hqfao.hub.arcgis.com/>. Last accessed 25 November 2021.
- Fuchs, R., Herold, M., Verburg, P.H., Clevers, J.G.P.W. (2013) A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in Europe, *Biogeosciences*, 10(3), 1543–1559, doi:10.5194/bg-10-1543-2013
- Fuchs, R., Verburg, P. H., Clevers, J.G.P.W., Herold, M. (2015) The potential of old maps and encyclopaedias for reconstructing historic continental land cover/use change. *Applied Geography*, 59, 43-55, doi:10.1016/j.apgeog.2015.02.013
- Gallego, F. J. (2010) A population density grid of the European Union. *Population and Environment*, 31(6), 460–473. doi:10.1007/s11111-010-0108-y
- Gallego, F. J., Batista, F., Rocha, C., and Mubareka, S. (2011) Disaggregating population density of the European Union with CORINE land cover. *International Journal of Geographical Information Science*, 25(12), 2051–2069. doi:10.1080/13658816.2011.583653
- Genest, C., Rémillard, B., Beaudoin, D. (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insur. Math. Econ.* 44, 199–213.
- Gisgraphy (2019) Download server, <http://download.gisgraphy.com/openstreetmap/pbf/>. Last accessed 23 July 2021.
- Global Dam Watch (2021) Global Reservoir and Dam Database (GRanD), <http://globaldamwatch.org/grand/>. Last accessed 7 July 2021.
- Gløersen, E. and Lüer, C. (2013) Population data collection for European local administrative units from 1960 onwards - final report. Spatial Foresight GmbH, Luxembourg, 17 pp. [data available at Eurostat, <http://ec.europa.eu/eurostat/web/nuts/local-administrative-units>]
- Goldsmith, R. W. (1985) Comparative national balance sheets: a study of twenty countries, 1688-1978. University of Chicago Press, Chicago, 364 pp.
- Groenemeijer, P., Vajda, A., Lehtonen, I., Kämäräinen, M., Venäläinen, A., Gregow, H., Becker, N., Nissen, K., Ulbrich, U., Paprotny, D., Morales Nápoles, O., and Púćik, T. (2016) Present and future probability of meteorological and hydrological hazards in Europe, D2.5

- report, RAIN project, available at http://rain-project.eu/wp-content/uploads/2016/09/D2.5_REPORT_final.pdf.
- Haupt, A., Kane, T. T., and Haub, C. (2011) Population Handbook (Sixth edition). Population Reference Bureau, Washington DC.
 - Hourihan, K. (1982) Urban Population Density Patterns and Change in Ireland, 1901-1979. *The Economic and Social Review*, 13(2), 125–147.
 - ILO and OECD (2015) The Labour Share in G20 Economies. Report prepared for the G20 Employment Working Group, available at <https://www.oecd.org/g20/topics/employment-and-social-policy/The-Labour-Share-in-G20-Economies.pdf>
 - International Monetary Fund (2021) World Economic Outlook Database, <http://www.imf.org/external/pubs/ft/weo/2017/01/weodata/index.aspx>. Last accessed 21 April 2021.
 - ISO (2015) Codes for the representation of currencies, ISO 4217:2015. Available at <https://www.currency-iso.org/en/home.html>
 - Joint Research Centre (2015) Global Land Cover 2000, <http://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php>. Last accessed 23 July 2021.
 - Klein Goldewijk, K., Beusen, A., and Janssen, P. (2010) Long term dynamic modeling of global population and built-up area in a spatially explicit way, HYDE 3.1. *The Holocene*, 20(4), 565–573. doi:10.1177/0959683609356587
 - Klein Goldewijk, K., Beusen, A., de Vos, M., and van Drecht, G. (2011) The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years. *Global Ecology and Biogeography*, 20(1), 73–86. doi:10.1111/j.1466-8238.2010.00587.x
 - Kummu, M., Taka, M., and Guillaume, J. H. A. (2018) Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Scientific Data*, 5, 180004. doi:10.1038/sdata.2018.4
 - Lehner, B., Reidy Liermann, C., Revenga, C., Vörösmarty, C., ..., Wisser, P. (2011) High resolution mapping of the world's reservoirs and dams for sustainable river flow management. *Frontiers in Ecology and the Environment*, 9(9), 494–502, doi:10.1890/100125
 - Li, Y., Rybski, D., Kropp, J.P. (2021) Singularity cities. *Environment and Planning B: Urban Analytics and City Science*, 48(1), 43-59, doi:10.1177/2399808319843534
 - Milego, R. and Ramos, M. J. (2011) Disaggregation of socioeconomic data into a regular grid and combination with other types of data. ESPON Technical Report, available at https://database.espon.eu/private-media/object/1886/report_Grid.pdf?ts=1640255791.5731616
 - Mitchell, B. R. (1998) International historical statistics: Europe 1750-1993. Macmillan, London
 - Murakami, D. and Yamagata, Y. (2019) Estimation of Gridded Population and GDP Scenarios with Spatially Explicit Statistical Downscaling. *Sustainability*, 11, 2106. doi:10.3390/su11072106
 - OECD (2021) OECD Data, <https://data.oecd.org/>
 - OpenStreetMap (2021), <https://www.openstreetmap.org/>. Last access 23 July 2021.
 - OurAirports (2021) Open data downloads, <http://ourairports.com/data/> . Last accessed 6 July 2021.
 - Papageorgiou, Y. Y. (2014) Population density in a central-place system. *Journal of Regional Science*, 54(3), 450–461. doi:10.1111/jors.12111
 - Paprotny, D. and Morales Nápoles, O. (2016) Pan-European data sets of river flood probability of occurrence under present and future climate, TU Delft, dataset, <http://dx.doi.org/10.4121/uuid:968098ce-afe1-4b21-a509-dedaf9bf4bd5>.

- Paprotny, D., Morales Nápoles, O., and Jonkman, S. N. (2017) Efficient pan-European river flood hazard modelling through a combination of statistical and physical models. *Natural Hazards and Earth System Sciences*, 17, 1267–1283, doi:10.5194/nhess-17-1267-2017.
- Paprotny D., Sebastian A., Morales Nápoles O., Jonkman S.N. 2018. Trends in flood losses in Europe over the past 150 years. *Nature Communications*, 9, 1985, doi:10.1038/s41467-018-04253-1.
- Paprotny, D., Morales Nápoles, O., Voudoukas, M. I., Jonkman, S. N., and Nikulin, G. (2019) Accuracy of pan-European coastal flood mapping. *Journal of Flood Risk Management*, 12(2), e12459, doi:10.1111/jfr3.12459.
- Paprotny, D., Kreibich, H., Morales-Nápoles, O., Terefenko, P., Schröter, K. (2020) Estimating exposure of residential assets to natural hazards in Europe using open data. *Natural Hazards and Earth System Sciences*, 20, 323–343, <https://doi.org/10.5194/nhess-20-323-2020>
- Piketty, T., and Zucman, G. (2014) Capital is Back: Wealth-Income Ratios in Rich Countries, 1700-2010. *Quarterly Journal of Economics*, 129(3), 1155-1210.
- Republic of North Macedonia State Statistical Office (2021) https://www.stat.gov.mk/Default_en.aspx
- Schmalwasser, O., Müller, A., and Weber, N. (2011) Gebrauchsvermögen privater Haushalte in Deutschland. *Wirtschaft und Statistik*, 06/2011, 565–578.
- Statistical Office of the Republic of Serbia (2021), <https://www.stat.gov.rs/en-us/>
- Taylor, B. (2004) Global History of Currencies, Global Financial Data, <https://www.globalfinancialdata.com/News/GHOC.aspx>
- United Nations (1974) Compendium of Housing Statistics 1971. Statistical Office of the United Nations, New York.
- United Nations (2015) Demographic Yearbook 2014. United Nations, New York.
- United Nations (2018) World Urbanization Prospects: The 2018 Revision, <https://population.un.org/wup/>
- United Nations (2019) The 2019 Revision of World Population Prospects, <https://population.un.org/wpp/>
- Wright, J.K. (1936) A Method of Mapping Densities of Population with Cape Cod as an Example. *Geographical Review*, 26(1), 103–110. doi:10.2307/209467

Appendix 1. Sources of administrative boundaries for NUTS region map

Table A1. Sources of administrative boundaries. Countries not listed here were collected from OpenStreetMap. All links accessed 5 May 2022.

Country	Provider	Dataset name	URL
Austria	Bundesamt für Eich- und Vermessungswesen	<i>Verwaltungsgrenzen (VGD) - Stichtagsdaten grundstücksgenau</i>	[1]
Belgium	FPS Finance - General Administration of Patrimonial Documentation (GAPD)	<i>Administrative units - Municipalities</i>	[2]
Estonia	Estonian Administrative and Settlement Division	<i>Counties</i>	[3]
Finland	The National Land Survey of Finland	<i>The regions of Finland 2021 with the 2018 regional codes [used in combination with OpenStreetMap data]</i>	[4]
Germany	Bundesamt für Kartographie und Geodäsie	<i>NUTS-Gebiete 1:250 000, Stand 01.01. (NUTS250 01.01.) [except Mecklenburg-</i>	[5]

		Vorpommern, where OpenStreetMap data were used]	
Greece	Hellenic Mapping and Cadastral Organisation	<i>Boundaries of Prefectures (HMCO); Boundaries of the Local Authorities (LAs) (pre-Kapodistrian)</i>	[6][7]
Ireland	Ordnance Survey Ireland	<i>Small Areas Ungeneralised - OSi National Statistical Boundaries – 2015; Local Electoral Areas Boundaries Ungeneralised - OSi National Administrative Boundaries - 2015</i>	[8][9]
Italy	Istituto Nazionale di Statistica (ISTAT)	<i>Municipal Boundaries of Italy 2019 [only for Sardegna, otherwise OpenStreetMap provinces used]</i>	[10]
Norway	Kartverket	<i>Administrative enheter kommuner</i>	[11]
Poland	Główny Urząd Geodezji i Kartografii	<i>PRG – jednostki administracyjne</i>	[12]
United Kingdom	Office for National Statistics	<i>Local Authority Districts (May 2020) Boundaries UK BFE</i>	[13]

[1] <https://www.data.gv.at/katalog/dataset/verwaltungsgrenzen-vgd-stichtagsdaten-grundstucksgenau>

[2] <https://www.geo.be/catalog/details/591e7f88-c443-4659-b8b7-23601d647ee6?l=en>

[3] <https://geoportaal.maaamet.ee/eng/Spatial-Data/Administrative-and-Settlement-Division-p312.html>

[4] https://www.avoindata.fi/data/en_GB/dataset/suomen-maakunnat-2021-vuoden-2018-maakuntakoodeilla

[5] <https://gdzshopv-lpz.bkg.bund.de/index.php/default/catalog/product/view/id/773/s/nuts-gebiete-1-250-000-stand-01-01-nuts250-01-01/category/8/?store=default>

[6] <http://geodata.gov.gr/en/dataset/oria-nomon-okkhe>

[7] <http://geodata.gov.gr/en/dataset/oria-ota-pro-kapodistria>

[8] <https://data-osi.opendata.arcgis.com/datasets/small-areas-ungeneralised-osi-national-statistical-boundaries-2015/explore?location=53.424600%2C-8.379100%2C7.63>

[9] https://data-osi.opendata.arcgis.com/datasets/1516dc49dfc64395b5a6ff582cba8150_3/explore?location=53.424600%2C-8.379100%2C7.63

[10] <https://hub.arcgis.com/datasets/inspire-esri::municipal-boundaries-of-italy-2019/about>

[11] <https://kartkatalog.geonorge.no/metadata/administrative-enheter-kommuner/041f1e6e-bdbc-4091-b48f-8a5990f3cc5b>

[12] <https://www.geoportal.gov.pl/dane/panstwowy-rejestr-granic>

[13] <https://geoportal.statistics.gov.uk/datasets/local-authority-districts-may-2020-boundaries-uk-bfe?geometry=-44.245%2C51.101%2C39.383%2C59.782>

Appendix 2. Detailed categories of non-financial assets

Table A2. Categories of non-financial assets included in, and excluded from, the study, according to ESA 2010 methodology. Items in red were excluded from the study. See European Union (2013), chapter 7, for detailed definitions and examples.

Code	Name
AN.1	Produced non-financial assets
AN.11	Fixed assets
AN.111	Dwellings
AN.112	Other buildings and structures
AN.1121	<i>Buildings other than dwellings</i>
AN.1122	<i>Other structures</i>
AN.1123	<i>Land improvements</i>
AN.113	Machinery and equipment
AN.1131	<i>Transport equipment</i>
AN.1132	<i>ICT equipment</i>
AN.1139	<i>Other machinery and equipment</i>
AN.114	Weapons systems
AN.115	Cultivated biological resources
AN.1151	<i>Animal resources yielding repeat products</i>
AN.1152	<i>Tree, crop and plant resources yielding repeat products</i>
AN.117	Intellectual property products
AN.12	Inventories
AN.13	Valuables
AN.2	Non-produced non-financial assets
AN.21	Natural resources
AN.211	Land
AN.212	Mineral and energy reserves
AN.213	Non-cultivated biological resources
AN.214	Water resources
AN.215	Other natural resources
AN.22	Contracts, leases and licences
AN.23	Purchases less sales of goodwill and marketing assets
AN.m	Consumer durables

Appendix 3. Local population estimates, 1961-2011

Gløersen and Luer (2013) collected population data at the level of local administrative units (LAUs) based on censuses made from ca. 1960 to 2011. The data were recalculated to a single set of subdivisions to allow consistent comparison through time. For most countries, data represent LAU level 2 (as defined by Eurostat), which is the lowest level of administrative divisions in a given country. For Lithuania and Slovenia a more aggregated level (LAU 1) was provided. The data was provided at census dates and interpolated (sometimes also extrapolated) to benchmark dates, spanning full decades from 1 January 1961 to 1 January 2011; the interpolated data were used in the analysis. The coverage of the study area is not complete. The following additions were made:

- In North Macedonia, there has been no census since 2002, hence mid-year estimates for 2011 from the Republic of North Macedonia State Statistical Office (2021) was used. Due to gaps in

data and spatial data limitations, the population figures were aggregated to LAU 1 level, with 10 LAUs in the Skopje area aggregated further to a single unit;

- Population of Serbia was taken from LAU 1 level data provided by the Statistical Office of the Republic of Serbia (2021);
- For five small countries and territories (Andorra, Isle of Man, Monaco, San Marino, the Vatican), population for the entire territory as compiled in the HANZE database was used.

The Eurostat's dataset doesn't also cover Albania, Bosnia and Herzegovina, Kosovo and Montenegro. Those countries were excluded from further analysis due to difficulty in gathering the necessary population and spatial data. Also, data is completely missing for 240 out of 614 LAUs in Cyprus, however the vast majority of those are located wholly or partially in Northern Cyprus, therefore outside this study's domain. Data is also missing for some years in 15 LAUs located in Czechia (3), Cyprus (1), France (1), Germany (3), Malta (2) and Slovakia (5).

The tabular data were merged with spatial data from Eurostat. As the subdivisions refer to different time points for different countries, and are also not always internally consistent, several datasets produced by EuroGeographics and provided to Eurostat were used ("Communes" for different years and "Census units 2011"). Due to the resolution of spatial data, population figures for Denmark, Greece, North Macedonia and Portugal had to be aggregated to LAU 1 level. For some countries, individual local units had to be aggregated, split or redrawn in order to match the subdivisions used in the population dataset. For 37 out of 3441 LAUs in Ireland and 86 out of 9499 LAUs in the United Kingdom it was possible to match tabular and spatial data due to large administrative boundary changes.

The final map contains 109,419 LAUs. Of these, 378 LAUs are not usable due to missing population data and 364 LAUs have no population¹⁵, resulting in a total of 108,679 LAUs with generally consistent population data for 1961-2011. It was noticed for a small number of cases (e.g. for some locations in Italy) that the data were not fully adjusted for administrative boundary changes. Still, the benefit of the dataset for analysing for population changes in Europe is very large.

¹⁵ These are non-municipal areas, e.g. lakes, forests or military bases, occurring in Cyprus, Estonia, France, Germany, Lithuania, North Macedonia, Spain and Switzerland.

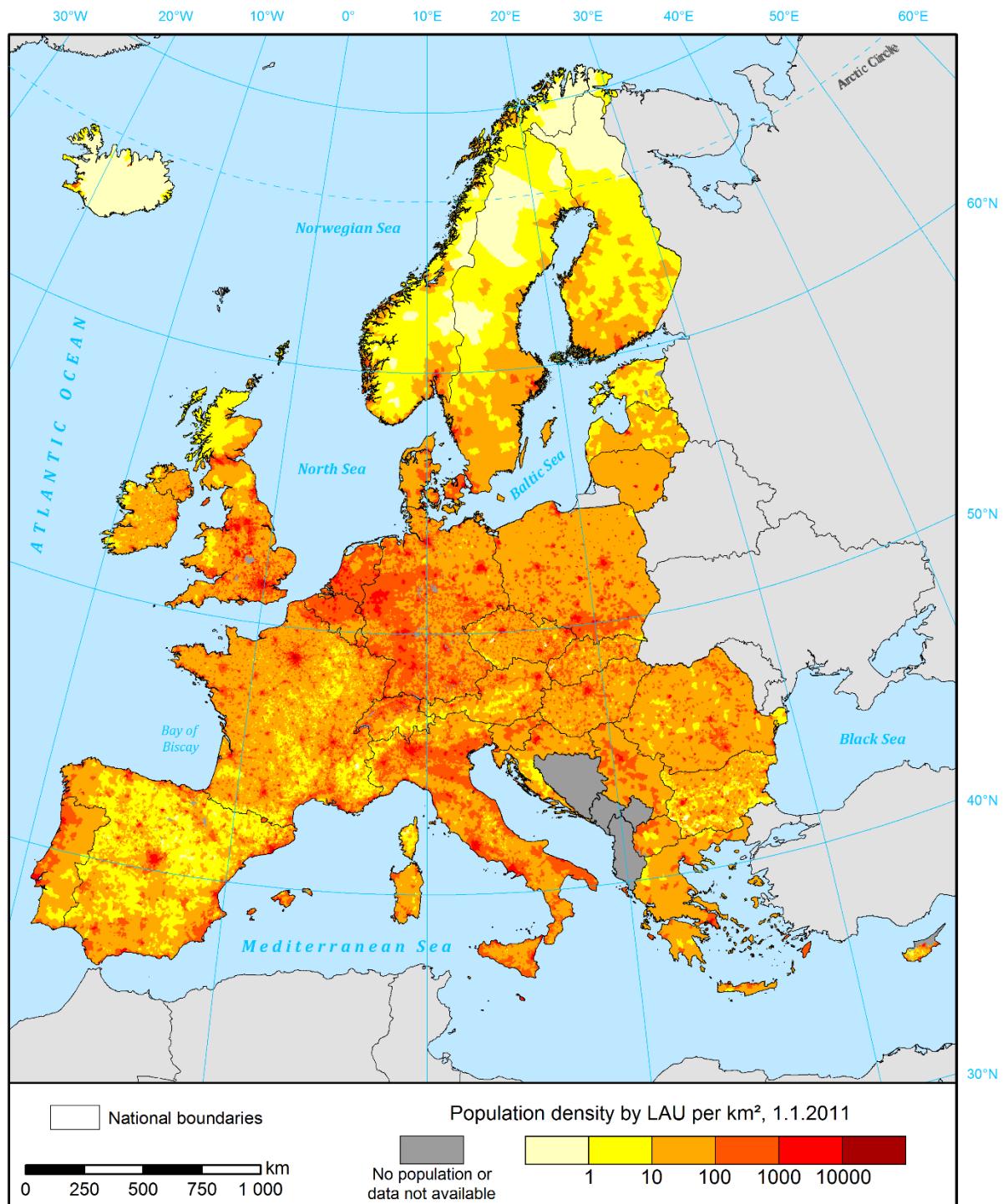


Figure A1. Population density per local administrative unit as of 2011 (see text for sources of data).

Appendix 4. Calibrating urban population change

The calibration utilizes Clark's (1951) model of population density, which he described with an exponential function:

$$y = Ae^{-bx} \quad (A1)$$

where y is the population density (in persons per ha), x is the distance from the city centre (in km), A and b are exponential function coefficients. Clark (1951, 1967) provided estimates of A and b for 16

cities in 9 countries for 29 time points. Hourihan (1982) provided additional estimates for 3 cities from several censuses, of which 13 cases were used (estimates made with only a few data points were excluded). That gives a total of 42 estimates spanning a whole century, from 1871 to 1971 (Table A3). In the population map constructed here the population density was calculated for 500 m wide zones around (arbitrarily chosen) city centre, interpolated to match the time points from literature and then fitted to an exponential function. The model was run many times using random weighting of five datasets indicating distance from urban centre. A comparison of function parameters for a most optimal combination is presented in Fig. A2. Overall, the fit is moderate, but a better match of modelled and observed estimates of eq. A1 parameters would be difficult, since the exponential curve fits are very sensitive to the sample size, i.e. spatial resolution of data and maximum distance from the city centre. The latter is not known for all cities. Additionally, the source literature studies used census wards of different sizes instead of a disaggregated population grid used here.

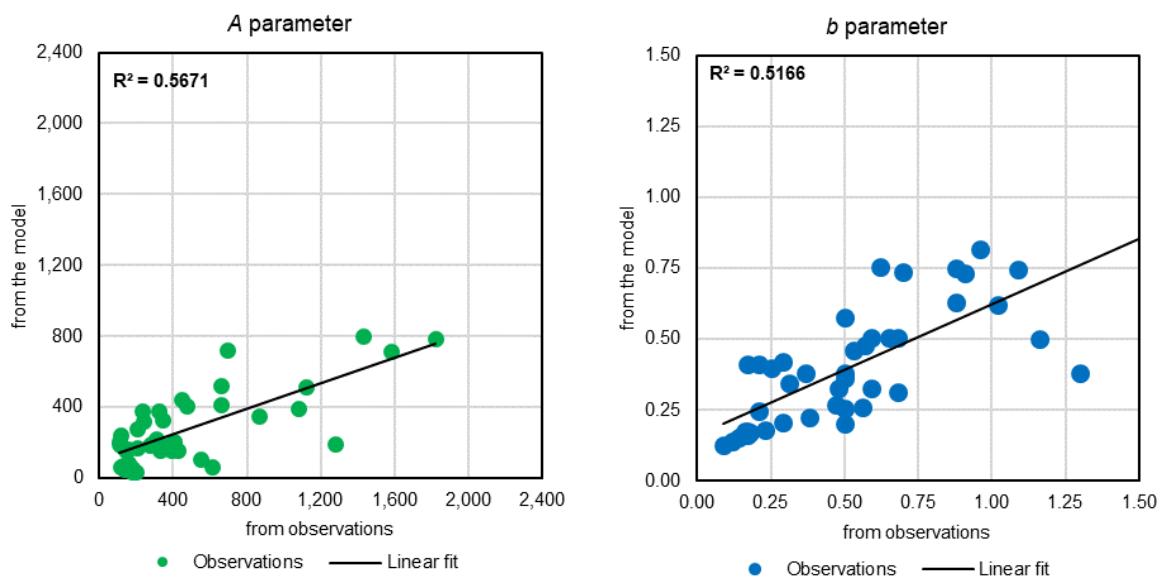


Figure A2. Estimates of A and b parameters (eq. A1) from modelled and observed population data.

Table A3. Estimates of urban population density. A, b – exponential function parameters (adjusted to give population density in persons per ha, rather than persons per sq. mile as in Clark 1951 and Hourihan 1982). D – maximum distance from the city centre (km), for which population data were used to calculate exponential function parameters (values in red are estimates, as the source does not specify the distance).

Name	Region	Year	A	b	D	Source
Aarhus	DK042	1950	279	0,96	5	Clark 1967
Berlin	DE300	1885	1120	0,68	8	Clark 1951; Clark 1967
Berlin	DE300	1900	1580	0,59	10	Clark 1951; Clark 1967
Birmingham	UKG31	1921	401	0,50	11	Clark 1967
Birmingham	UKG31	1938	201	0,29	12	Clark 1967
Budapest	HU101	1935	1080	0,56	5	Clark 1951; Clark 1967
Copenhagen	DK011	1940	231	0,37	7	Clark 1967
Cork	IE025	1926	199	1,02	3	Hourihan 1982
Cork	IE025	1936	177	0,88	3	Hourihan 1982
Cork	IE025	1951	176	0,91	4	Hourihan 1982
Cork	IE025	1961	114	0,70	4	Hourihan 1982

Cork	IE025	1971	158	0,62	4	Hourihan 1982
Dublin	IE021	1901	391	0,68	4	Hourihan 1982
Dublin	IE021	1911	379	0,65	4	Hourihan 1982
Dublin	IE021	1926	352	0,59	4	Hourihan 1982
Dublin	IE021	1936	270	0,53	6	Clark 1951; Clark 1967
Dublin	IE021	1951	106	0,25	8	Hourihan 1982
Dublin	IE021	1961	105	0,21	8	Hourihan 1982
Dublin	IE021	1971	113	0,17	8	Hourihan 1982
Frankfurt am Main	DE712	1890	550	1,16	5	Clark 1967
Frankfurt am Main	DE712	1933	340	0,57	6	Clark 1967
Leeds	UKE42	1951	116	0,31	9	Clark 1967
Limerick	IE023	1961	136	1,09	3	Hourihan 1982
Limerick	IE023	1971	126	0,88	3	Hourihan 1982
Liverpool	UKD72	1921	1275	0,50	9	Clark 1951; Clark 1967
London	UKI11	1871	865	0,38	17	Clark 1967
London	UKI11	1901	660	0,23	20	Clark 1967
London	UKI11	1921	443	0,17	25	Clark 1967
London	UKI11	1931	475	0,17	26	Clark 1967
London	UKI11	1939	320	0,14	28	Clark 1967
London	UKI11	1951	240	0,12	29	Clark 1967
London	UKI11	1961	205	0,09	33	Clark 1967
Manchester	UKD31	1931	155	0,16	18	Clark 1951
Manchester	UKD31	1939	143	0,18	20	Clark 1967
Oslo	NO011	1938	308	0,50	4	Clark 1951; Clark 1967
Paris	FR101	1896	1430	0,50	12	Clark 1951; Clark 1967
Paris	FR101	1931	1820	0,47	14	Clark 1951; Clark 1967
Paris	FR101	1946	695	0,21	16	Clark 1967
Stockholm	SE110	1880	610	1,30	10	Clark 1967
Stockholm	SE110	1940	425	0,48	10	Clark 1967
Vienna	AT130	1890	660	0,50	7	Clark 1951; Clark 1967
Zurich	CH040	1936	328	0,29	8	Clark 1967

Appendix 5. Data for the land-use transition model

Data for the land-use transition model, implemented as Bayesian Network, was obtained by sampling the CLC inventory. Firstly, vector layers of CLC-Changes 2000-2006, 2006-2012 and 2012-2018 were obtained (1,194,980 patches). The CLC classes before and after transition were grouped together as follows:

- Urban fabric (CLC 111-112);
- Other artificial (CLC 121-142);
- Croplands (CLC 211-223 and 241-244);
- Pastures (CLC 231);
- Natural (CLC 311-324, 333 and 411-412);
- Other (CLC 331-332, 334-335, 421-523).

CLC-Changes patches transitioning within a given group, or transitioning from/to “Other” classes were excluded from further analysis, leaving 240,870 patches of varying size. To sample the inventory, a “fishnet” of 25x25 km cells was created and clipped to the land mask of the baseline land cover/use map. It was further clipped to the remaining CLC-Changes patches. In each cell of the fishnet, $1 + \frac{A}{7}$ samples (rounded to the nearest integer) within the borders of CLC-Changes patches were generated¹⁶, where A is the area of all patches within a fishnet cell (in hectares). The samples were generated at least 100 meters apart due to the resolution of most raster data used throughout this study. In total, 513,915 locations with land-use transitions were obtained. Then, an equal number of samples in each fishnet cell were generated in the remaining area of the cells. In this way, the same number of random locations where no land-use transitions took place between 2000 and 2018 was obtained with the same spatial distribution as the other dataset. Instances of transitions are many times fewer in reality, but as we are interested in the relative probability of transition between CLC classes rather than the total probability, we can sample a greater proportion of transitions to better quantify the patterns of land-use changes. An example of sample locations is presented in Fig. A3.

A smaller validation dataset was created as well. A new set of random points was generated in the fishnet, but the number of samples per cell was capped at 15, resulting in 97,790 samples each for transitions and non-transitions, with a much more even spatial distribution throughout Europe.

The sample locations were used to extract data from various raster datasets, as follows:

- Population density in 2011 computed using kernel density and per “virtual” LAU (see section 3.4.1);
- Euclidean distances from urban centres based on five datasets (see section 3.4.1);
- Elevation (in meters) and slope (per mille) computed from EU-DEM dataset at 100 m resolution (Eurostat 2021b);
- Agricultural suitability indices calculated by FAO (2021b) in the Global Agro-Ecological Zoning version 4 (GAEZ) database for five different crops (alfalfa, grass, wheat, rye and white potato):
 - Suitability index range (0–10000), with all land in grid cell under rainfed conditions;
 - Output density (potential production divided by total grid cell area) under rainfed conditions;
 - Agro-climatic potential yield with an available water content of 200 mm/m (under irrigation conditions).

The agricultural indices were all for the historical period 1971-2000 using CRUTS32 climate data, assuming high input level and without CO₂ fertilization. The indices combine data on climate, soil and terrain to estimate potential yield of various crops. The resolution of this dataset is 5' (about 4–7 km, depending on location).

¹⁶ Using „Create Random Points” tool in ArcGIS 10.7.1

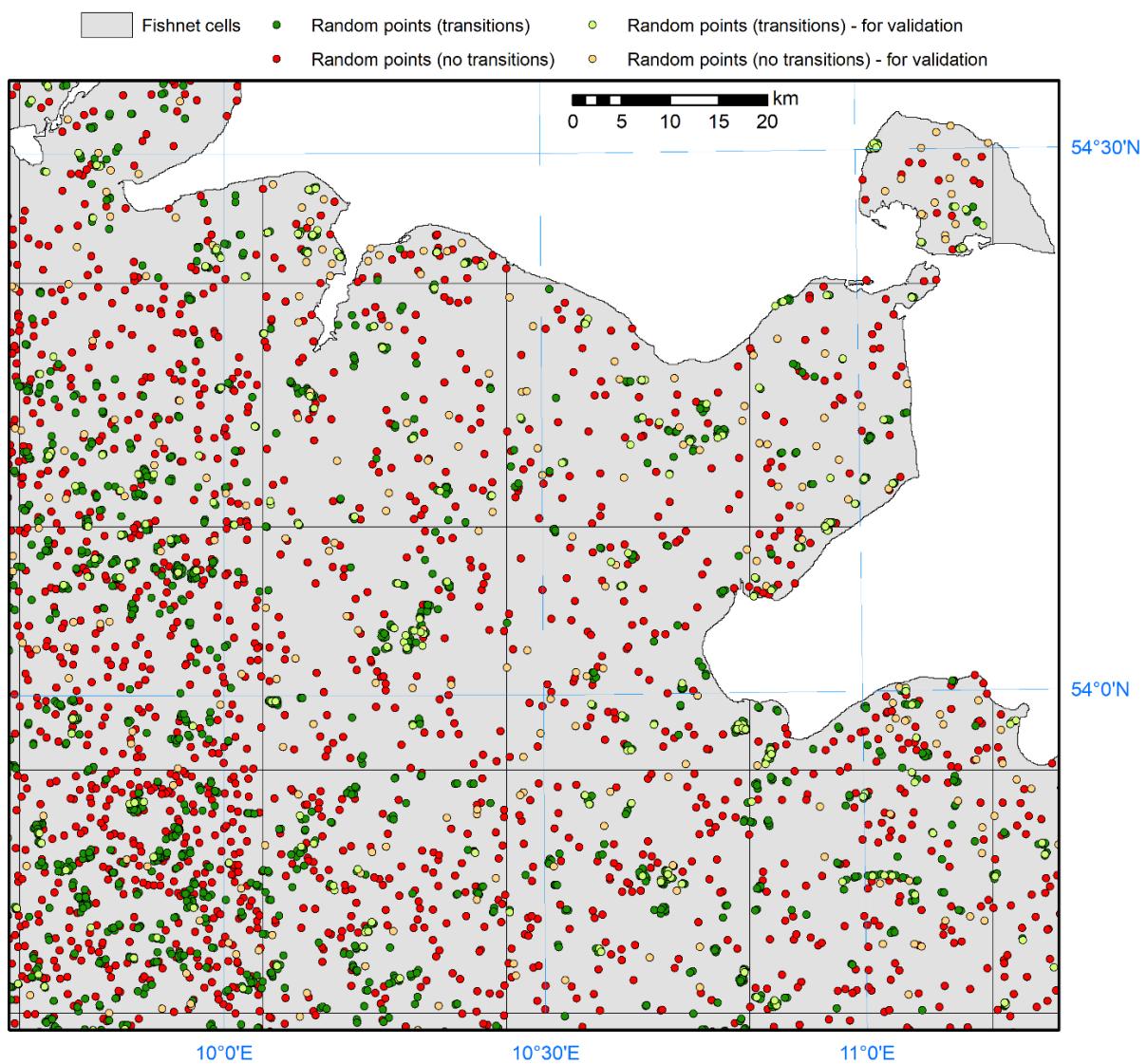


Figure A3. Example of sample locations generated for a fragment of northern Germany.

Appendix 6. Supplementary maps

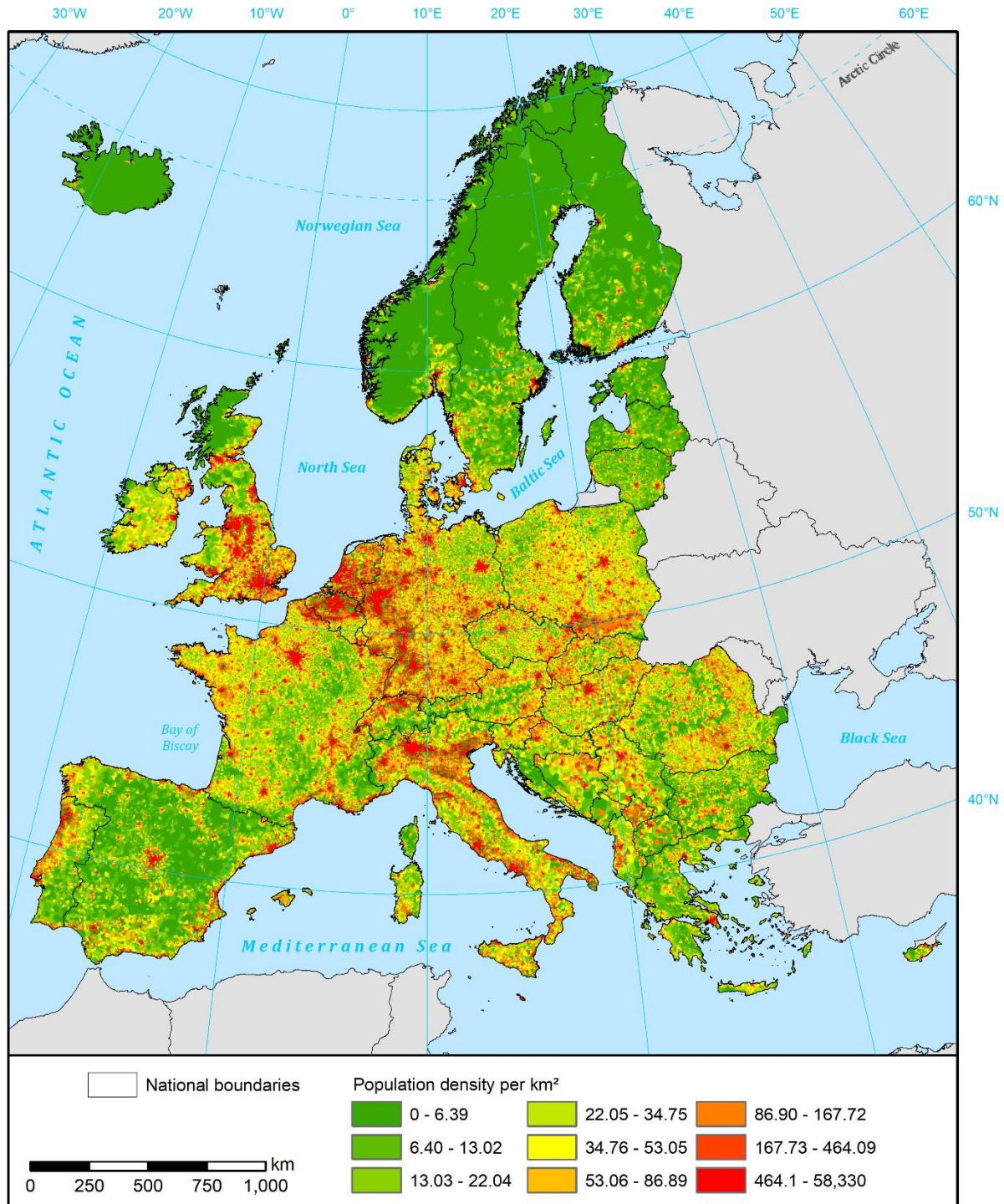


Figure A4. Population density by “virtual LAU” divided into 9 bins.

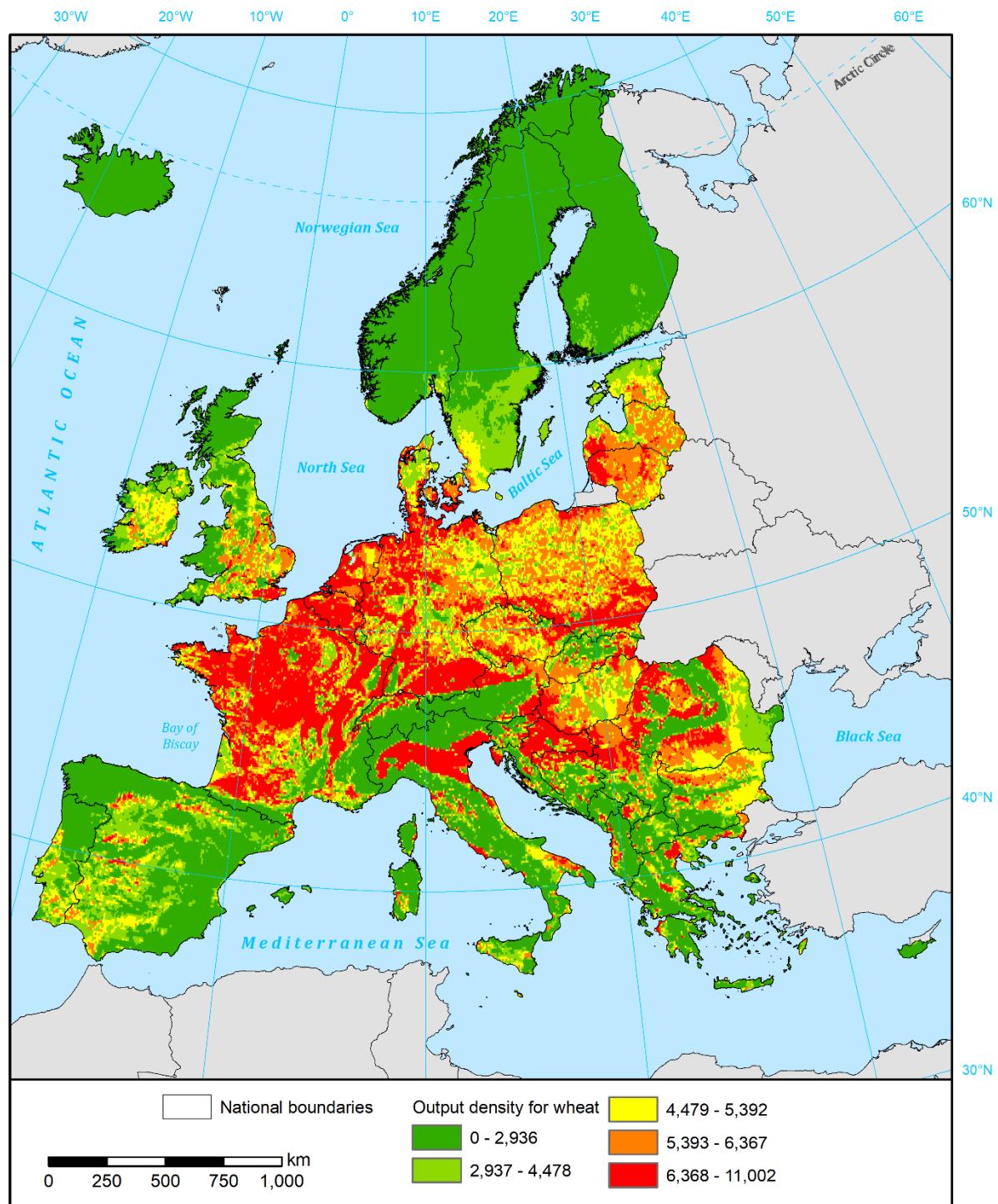


Figure A5. Output density (potential production divided by total grid cell area) for wheat under rainfed conditions and high input level for 1971-2000 from FAO's Global Agro-Ecological Zoning version 4 database.

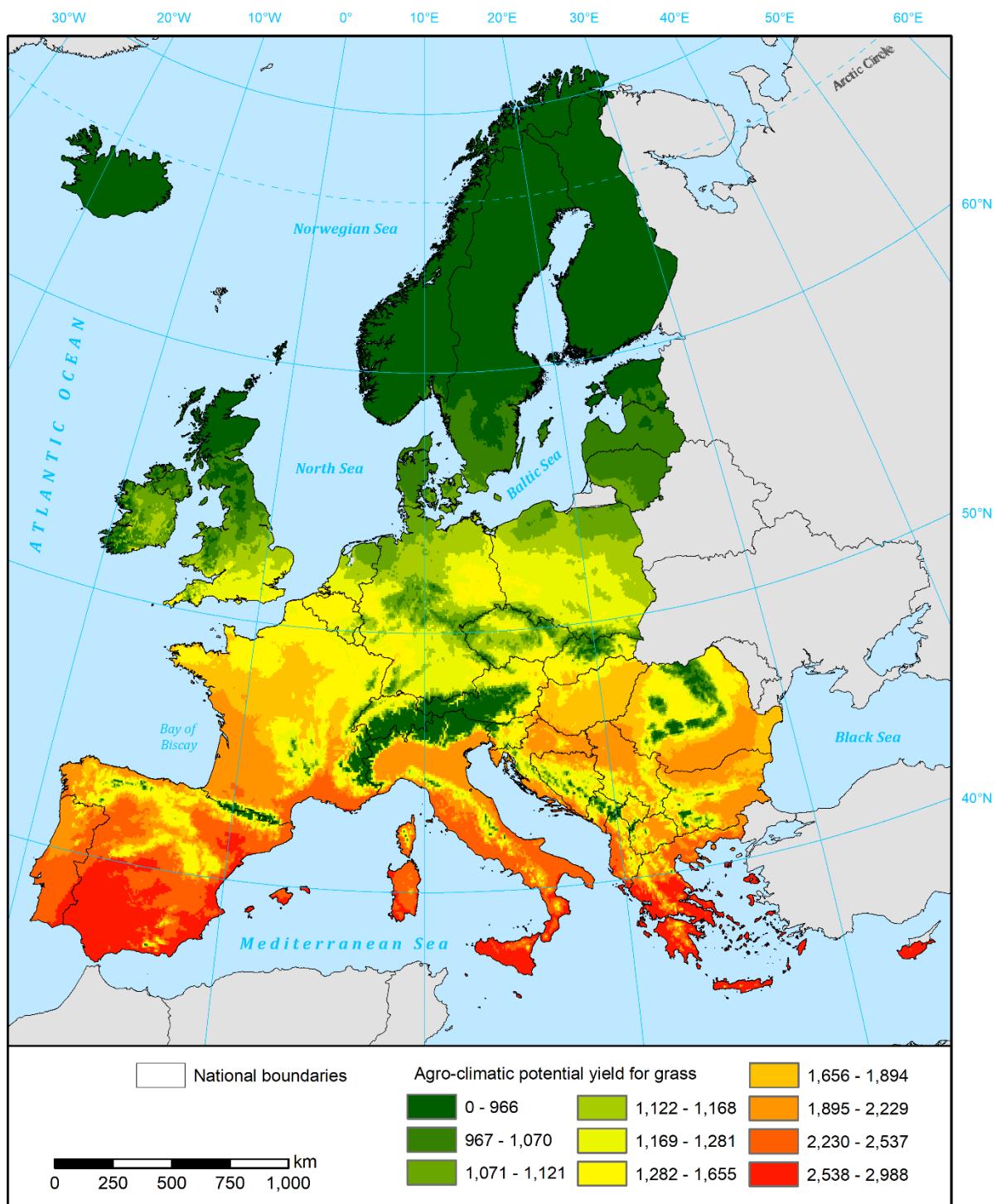


Figure A6. Agro-climatic potential yield for grass with an available water content of 200 mm/m (under irrigation conditions) and high input level for 1971-2000 from FAO's Global Agro-Ecological Zoning version 4 database.

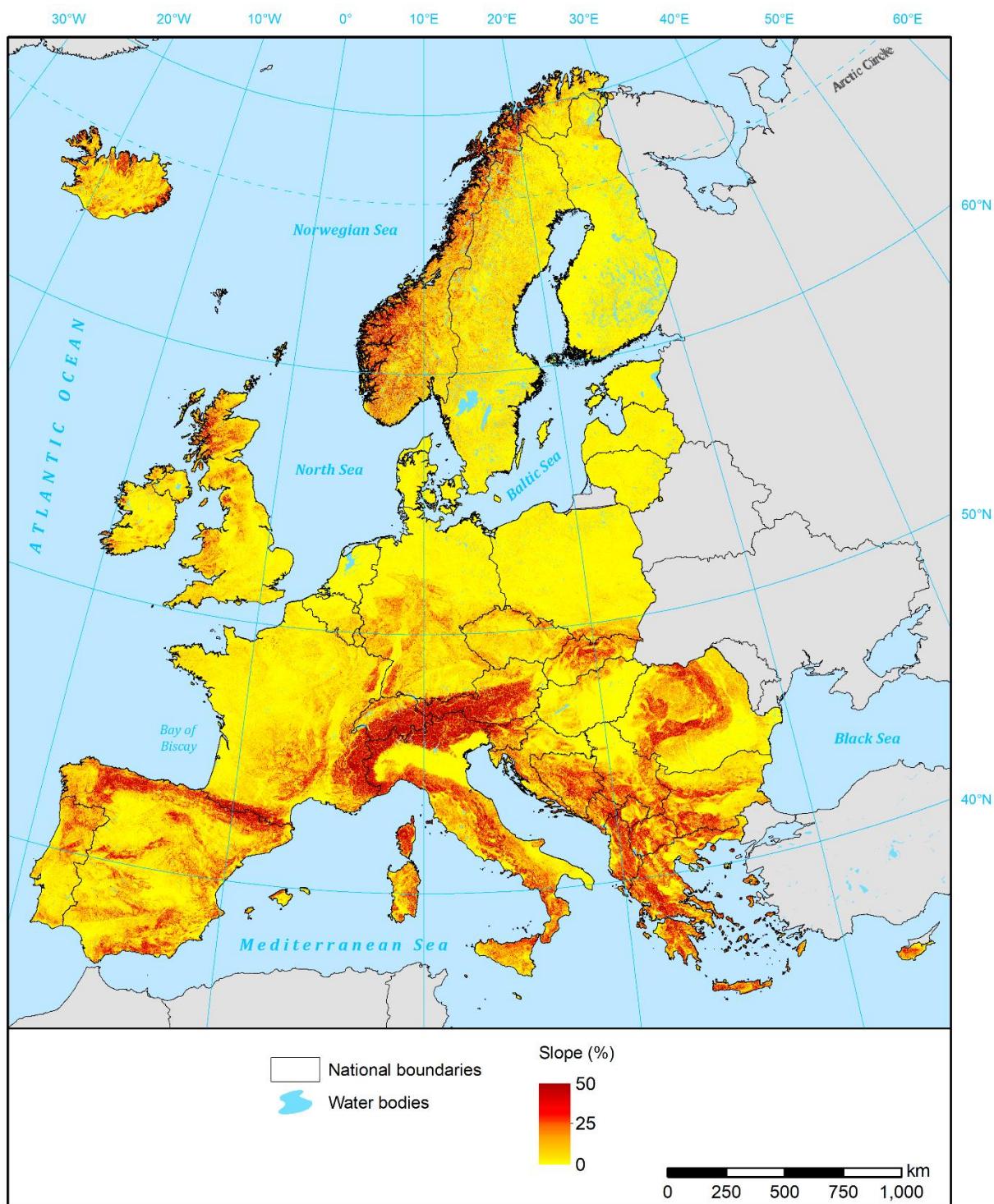


Figure A7. Slopeness of terrain in Europe, according to EU-DEM.