HANZE: Historical Analysis of Natural Hazards in Europe

Database documentation

Dominik Paprotny, Oswaldo Morales-Nápoles, Sebastiaan N. Jonkman

Department of Hydraulic Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands

Correspondence to: Dominik Paprotny (d.paprotny@tudelft.nl)

Document version: 1.1, 20 March 2018.

Summary. HANZE database, or 'Historical Analysis of Natural Hazards in Europe', aims to provide information on exposure to natural hazards for 37 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 1564 events (1870–2016). Demographic and economic data encompassed in HANZE allow to 'normalize' information on losses due to natural hazards by taking into account 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.

Table of Contents

Citation

Earth System Science Data description paper:

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.

This documentation:

Paprotny, D., Morales-Nápoles, O., Jonkman, S. N. (2017) HANZE: Historical Analysis of Natural Hazards in Europe – database documentation. TU Delft, Delft, the Netherlands.

The HANZE dataset:

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 (2017).

Acknowledgements

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 Ha**z**ards 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.

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.

2.2. Coverage

2.2.1. Spatial coverage

HANZE covers most of the European continent. Included are all 28 European Union member states, all four European Free Trade Agreement members (Iceland, Liechtenstein, Norway and Switzerland), four microstates located in Western Europe (Andorra, Monaco, San Marino and the Vatican) and one Crown Dependency of the United Kingdom (Isle of Man). Excluded are, therefore, non-EU successor states of the Soviet Union or Yugoslavia, as well as Albania and 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.

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

Figure 1. HANZE domain.

2.2.2. Temporal coverage

HANZE-Exposure covers the period of 1870–2020, with data for 1870–1970 having a 10-year time step, and starting with 1975 a 5-year time step. The exception are some economic data (see Table 1), which are provided annually. 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 II. Such situations are noted in the metadata of the database.

HANZE-Events covers flood events that occurred during the period of 1870–2016, with daily dates of the beginning and end of the events represented in the database.

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, 5/10-yearly	3.2.1, 3.3.6, 3.5
Output	Total population	Gridded 100 m, 5/10-yearly	3.2.2, 3.3.3, 3.5
Output	Gross domestic product (GDP) per year (euro)	Gridded 100 m, 5/10-yearly	3.3.7, 3.6
Output	Value of wealth	Gridded 100 m, 5/10-yearly	3.4.1, 3.6
Input	Total population	Regions, 5/10-yearly	3.3.3
Input	Urban population (% of total population)	Regions, 5/10-yearly	3.3.4
Input	Mean number of persons per household	Regions, 5/10-yearly	3.3.5
Input	Land use structure, selected types	Regions, 5/10-yearly	3.3.6
Input	GDP per year (euro)	Regions, 5/10-yearly	3.3.7
Input	GDP structure (% of GDP)	Regions, 5/10-yearly	3.3.7
Input	Value of wealth by category (% of GDP)	Countries, 5/10-yearly	3.3.8
Input	Forestry (% of agricultural sector GDP)	Countries, 2011	3.3.7
Auxiliary	GDP deflator	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 3.3.1). 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 project[1]. 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 cacluation for the 1953 coastal flood in the Netherlands is presented in Table 2. In the affected regions' 100-year coastal flood zone[2], according to HANZE-Exposure, population increased by 60% between`1953 and 2011, while GDP increased 5.6 times and wealth 7.4 times. 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 9153 coastal flood in the Netherlands.

Category	Reported losses (1953)	Exposure (1953)	Exposure (2011)	Relative losses	Normalized losses (2011)
Area flooded	2000	3917	3917	51.1%	2000
Persons killed	1835	1,245,000	1,988,000	0.15%	2930
Persons affected	188,000	1,245,000	1,988,000	15.1%	300,100
Losses in mln euro (2011 prices)	4.8 bln	13.6 bln 46.5 bln	75.8 bln 341.8 bln	35.5% 10.4%	26.9 bln 35.5 bln

* upper figure – GDP, lower figure – wealth.

2.4. List of files

A list of files with explanations is shown in Table 3.

All raster files are in ETRS89 / LAEA projection (EPSG:3035) and have a resolution of 100 m.

Table 3. List of files of HANZE database. XXXX = value indicating the year to which dataset pertains.

File	Format	Variables / contents
Output		
CLC_XXXX	8-bit TIFF raster	Land cover/use type, 44 classes according to Corine Land Cover (section 3.2.1)
Pop_XXXX	16-bit TIFF raster	Total population per grid cell (in persons)
GDP_XXXX	16-bit TIFF raster	Gross domestic product (GDP) per grid cell per year (x 10,000 euro in constant 2011 prices)
FA_XXXX	16-bit TIFF raster	Wealth per grid cell (x 100,000 euro in constant 2011 prices)
Exposure_5min	netCDF	Land use (fraction of: urban areas, croplands, pastures, forests, water), total population, GDP and wealth per grid cell in 5 arc minute resolution

[1] Data downloadable from Paprotny and Morales-Napoles (2016). For general description of the data, see Groenemeijer et al. (2016), and for details we refer to Paprotny et al. (2016, 2017).
[2] The affected regions are: Groot-Rijnmond (NL339), Zuidoost-Zuid-Holland (NL33A), Zeeuwsch-Vlaanderen (NL341), Overig Zeeland (NL342), West-Noord-Brabant (NL411).

Exposure_cordex _0.11	netCDF	Land use (fraction of: urban areas, croplands, pastures, forests, water), total population, GDP and wealth per grid cell in EURO-CORDEX rotated grid, 0.11° resolution.
Events_floods	Excel file	List of past damaging floods (list of variables in section 4.2)
Input / Auxiliary		
Expo_input_CLC_Pop	Excel file	Input land use/cover and population data (list of variables and table structures in section 3.3.1)
Expo_input_Econ	Excel file	Input and auxiliary economic data (list of variables and table structures in section 3.3.1)
CLC_base	8-bit TIFF raster	Baseline land cover/use type, 44 classes according to Corine Land Cover (section 3.2.1)
Pop_base	16-bit TIFF raster	Total baseline (disaggregated) population per 100 m grid cell (in persons)

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 is compiled for one 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, outlying preparation of baseline maps (section 3.2), compiling a database of regional-level statistical data (3.3), modelling of changes in land use and population distribution (3.4), disaggregation of economic variables (3.5) and production of final exposure maps with a 100 m resolution (3.6).

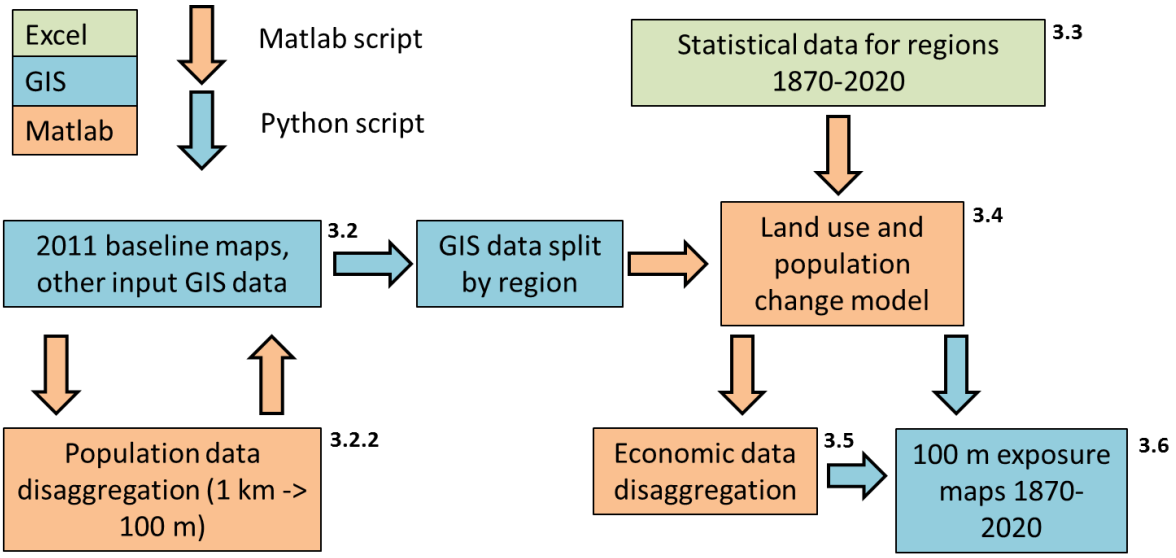

Figure 2. Workflow of HANZE-Exposure. Numbers next to the boxes refer to sections in this document.

3.2. Baseline maps

3.2.1. Land cover/use

The baseline land cover/use is based on Corine Land Cover (CLC) 2012, version 18.5a (Copernicus 2017). 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 and 2012. The maps are prepared, in general, by manual classification of land cover patches from satellite imagery. For the latest 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 was found to be higher than 85% (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 v18.5a, which covers a small strip around the border;

2) CLC 2000 v18.5, which covers a larger strip around the border;

3) Open Street Map from Gisgraphy (2016);

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.

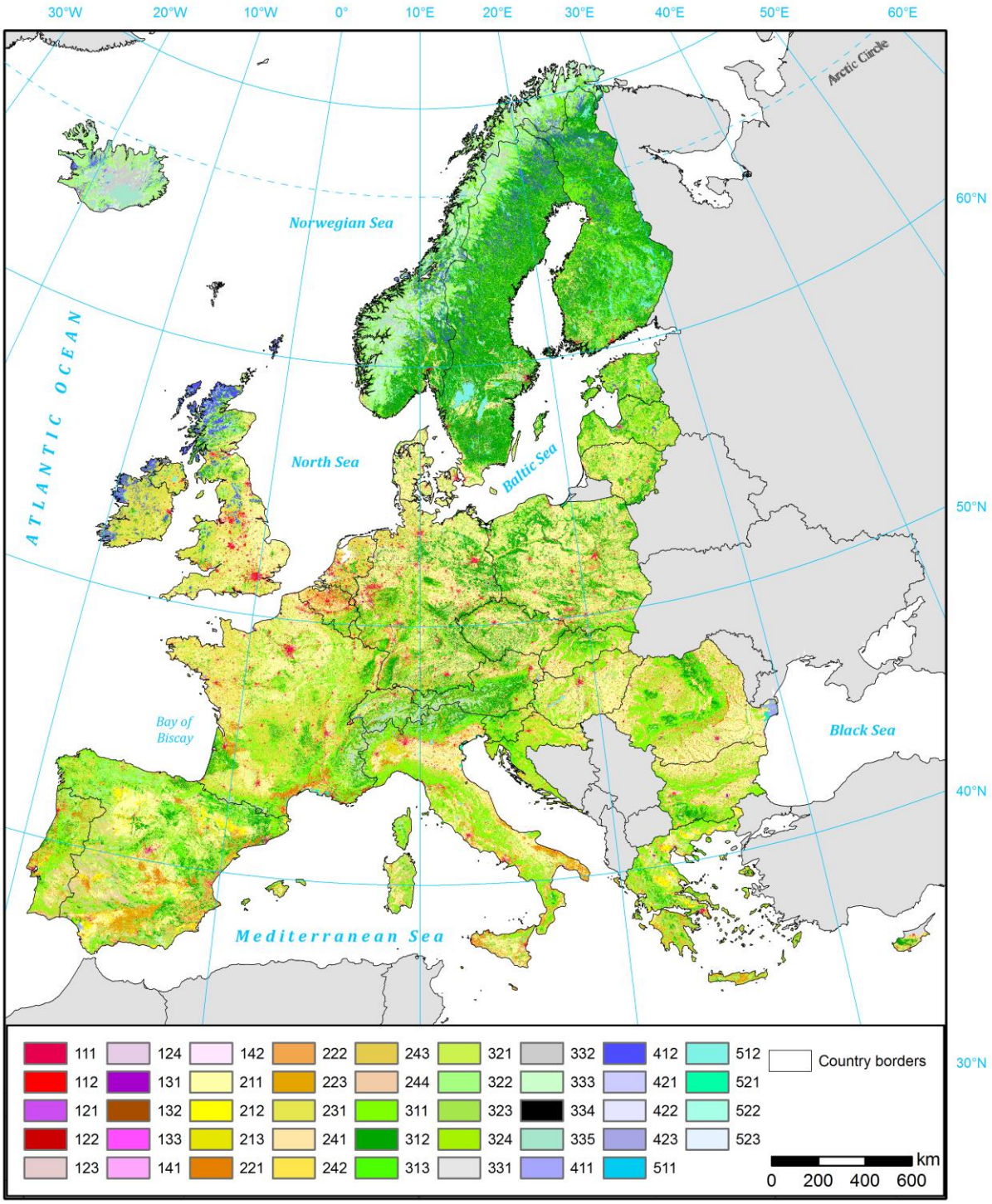

Figure 3. Baseline land cover, 2011, based on Corine Land Cover 2012. For explanation of CLC classes, see Fig. 3.

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

3.2.2. Population

The baseline population map is based on GEOSTAT population grid for the year 2011, version 2.0.1 (Eurostat 2016b). 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, Cyprus, Iceland, Isle of Man, Luxembourg, Monaco, San Marino 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.

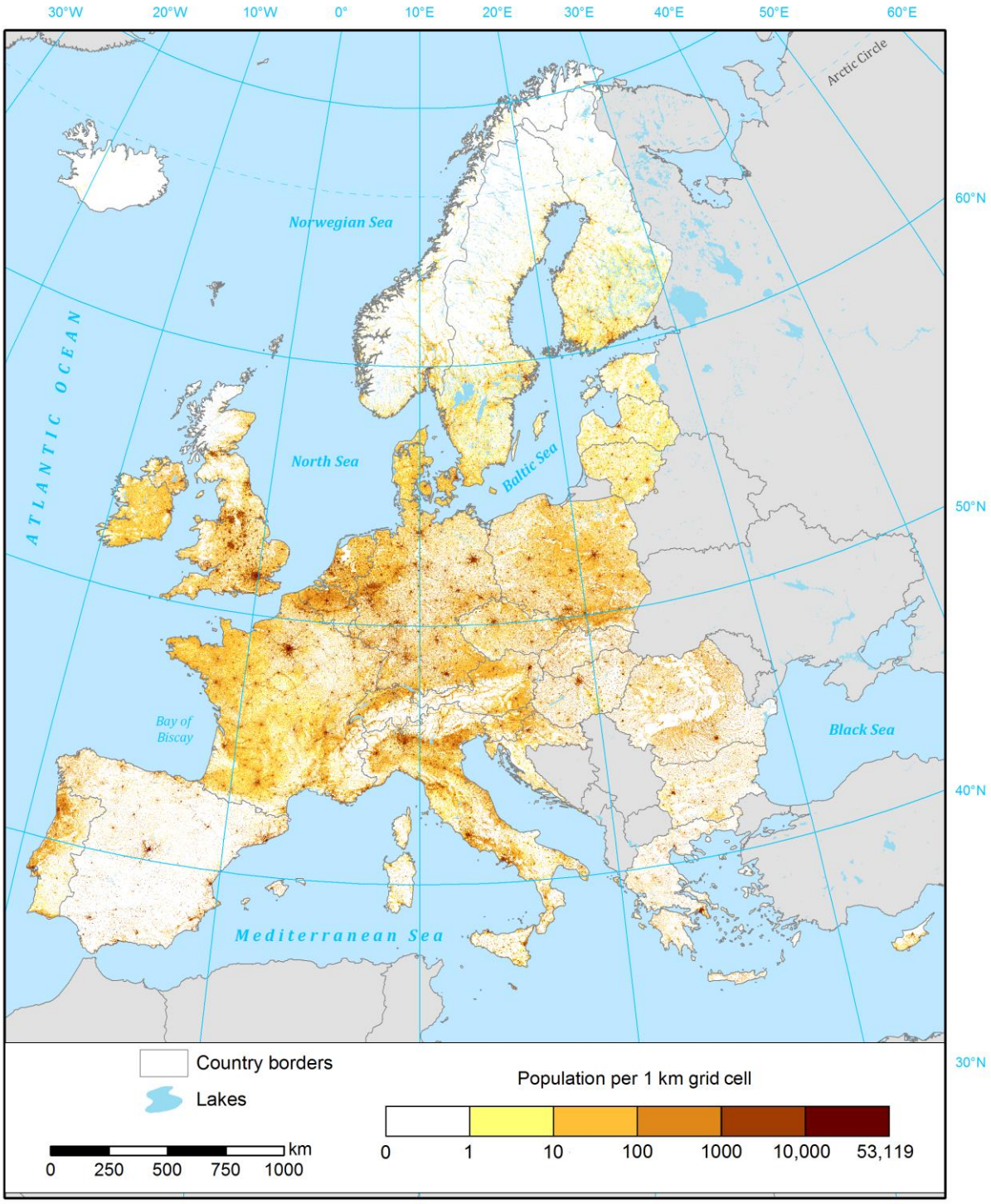

Figure 5. GEOSTAT population grid, 1 km resolution.

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 population 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}:

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

- 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 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.

Table 4. Thresholds for population disaggregation algorithm $\mathbf{T_L}$

CLC class name and code	Threshold (persons per km^2)
Continuous urban fabric (111)	22666
Discontinuous urban fabric (112)	6452
Other artificial (121–142)	59
Non-irrigated arable land (211)	32
Permanently irrigated land (212)	64
Rice fields (213)	9
Vineyards (221)	50

Fruit trees and berry plantations (222)	44
Olive groves (223)	60
Pastures (231)	40
Annual crops associated with permanent crops (241)	71
Complex cultivation patterns (242)	82
Land principally occupied by agriculture (243)	40
Agro-forestry areas (244)	10
Broad-leaved forest (311)	9
Coniferous forest (312)	6
Mixed forest (313)	9
Natural grasslands (321)	18
Moors and heathland (322)	18
Sclerophyllous vegetation (323)	10
Transitional woodland-shrub (324)	11
Sparsely vegetated areas (333)	40
Uninhabitable natural areas (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 level of soil sealing, or imperviousness of the ground. This variable has a range from 0%, which indicates completely natural surface, and 100%, which indicates land completely sealed by an artificial surface. This information could not be used directly to redistribute the population as large soil sealing may be caused both by residential and non-residential buildings as well as infrastructure. However, large elements of infrastructure or industry were already taken into account using the 'limiting variable' method.

Data on soil sealing were obtained from the Imperviousness 2012 dataset from Copernicus (2017). It was created based on high-resolution satellite photos taken during 2011-12 in visible and infrared spectrum. This dataset has a 100-meter resolution, which was resampled to a 1 km grid, so that average population density in grid cells with given imperviousness could be calculated. The resulting relationship can be approximated as a power function, based on cells imperviousness ranging from 1% to 96%, as very few cells have values above 96% (Fig. 6). Hence, the population X_g in 100-meter grid cell *g* is equal to:

$$X_g = \frac{Z_g}{\sum Z_g} Y_{LG} S_{LG} \tag{6}$$

where Z_g is the population of grid cell *g* obtained from the power function divided by maximum population (at 96% imperviousness):

$$Z_g = \frac{19.479 V_g^{1.3195}}{8031} \tag{7}$$

where V_g is the imperviousness in grid cell *g*. The population X_g is rounded, as population numbers need to be integers. However, rounding can cause difference between the population X_{LG} before and after disaggregation through soil sealing. In such a case, the population is added or subtracted randomly (with equal probability) within the land use class, one persons at the time, until the population X_{LG} matches the value before the second stage of disaggregation.

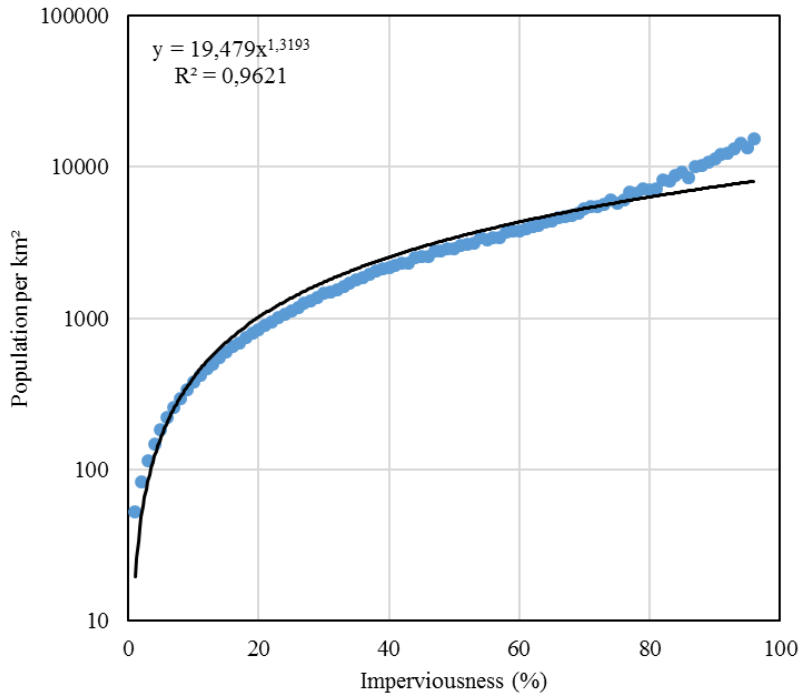

Figure 6. Relationship between average population density per imperviousness (soil sealing) class.

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 Delft, in the Netherlands. In this grid cell, the population at the time of the 2011 census was 1218. The top left box is an extract from the 1:25,000 topographic map. The top right box shows the land use structure according to Corine Land Cover 2012, and the bottom right box shows soil sealing according to Imperviousness 2012 dataset. The final 100 m population grid, based on aforementioned disaggregation process, is presented in the bottom left box.

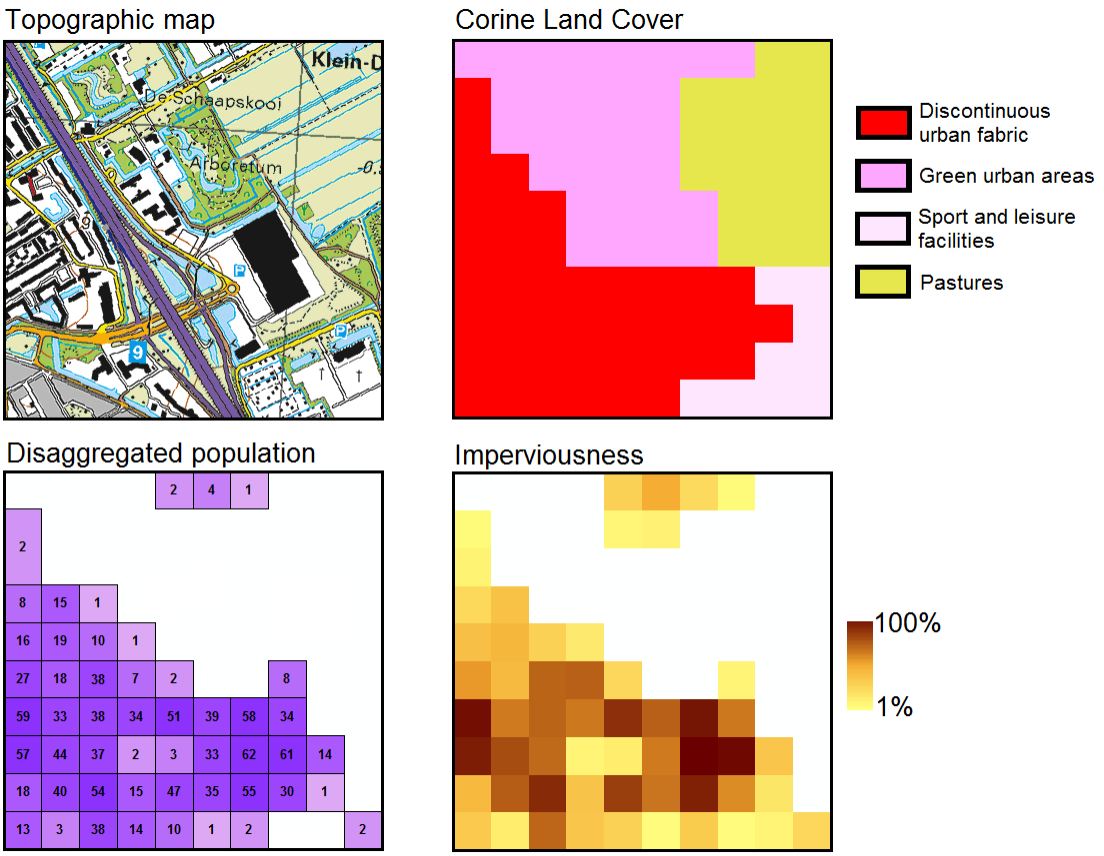

Figure 7. Disaggregation result and source data (population in the grid = 1218), contrasted with a topographic map. Fragment of city of Delft, The Netherlands (region NL333 'Delft en Westland').

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 5 and 6).

Population and land cover/use - Expo_input_CLC_Pop

Table 5. 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	Code – NUTS0 country code

	the 2011 censuses, which are the baseline population figures	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	CLC2012 – Corine Land Cover 2012 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	CLC2012 – Corine Land Cover 2012 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 6. Contents of economic data file - Expo_input_Econ

Variable	**Unit**	**Table structure**
GDP	Million euro in constant 2011 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	Code – NUTS0 country code Name – country/territory name 1870…2020 – data by year
Wealth in agriculture	% of GDP from agriculture	
Wealth in industry	% of GDP from industry	
Wealth in services	% of GDP from services	
Wealth in infrastructure	% of total GDP	
Forestry index	% of GDP from agriculture	Code – NUTS0 country code Name – country/territory name Index – 2011 share of forestry in agricultural GDP
Deflator	Index, 1990 or 2011 = 100	Code – NUTS0 country code Name – country/territory name 1870…2020 – data by year Unit – unit of measure (2011 = 100 or 1990 = 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,	Country – NUTS0 country code Currency – currency code Code – merged NUTS0 and

	2011 exchange rates were used.	currency code Conversion – conversion factor

Both files also contain:

- Sources, which explain the sources of data, transformations made to the original data and methods to estimate gaps in the data, dive
- References, which lists all publications mentioned in “Sources”.

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. A vector map of regions was obtained from ESRI (2016) with amendments based on Eurostat (2016a) in order to fully match NUTS 2010 classification. Coastlines in the vector map were further adjusted using Corine Land Cover 2012 map. 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. In the database there is a total of 1353 NUTS 3 regions (Fig. 8). A region has an average area of 3580 km^2 and an average total population of 379,000 as of 2011 census. Almost a third of all regions are located in Germany (412), since they are typically smaller than in most other countries (average population is only 195,000).

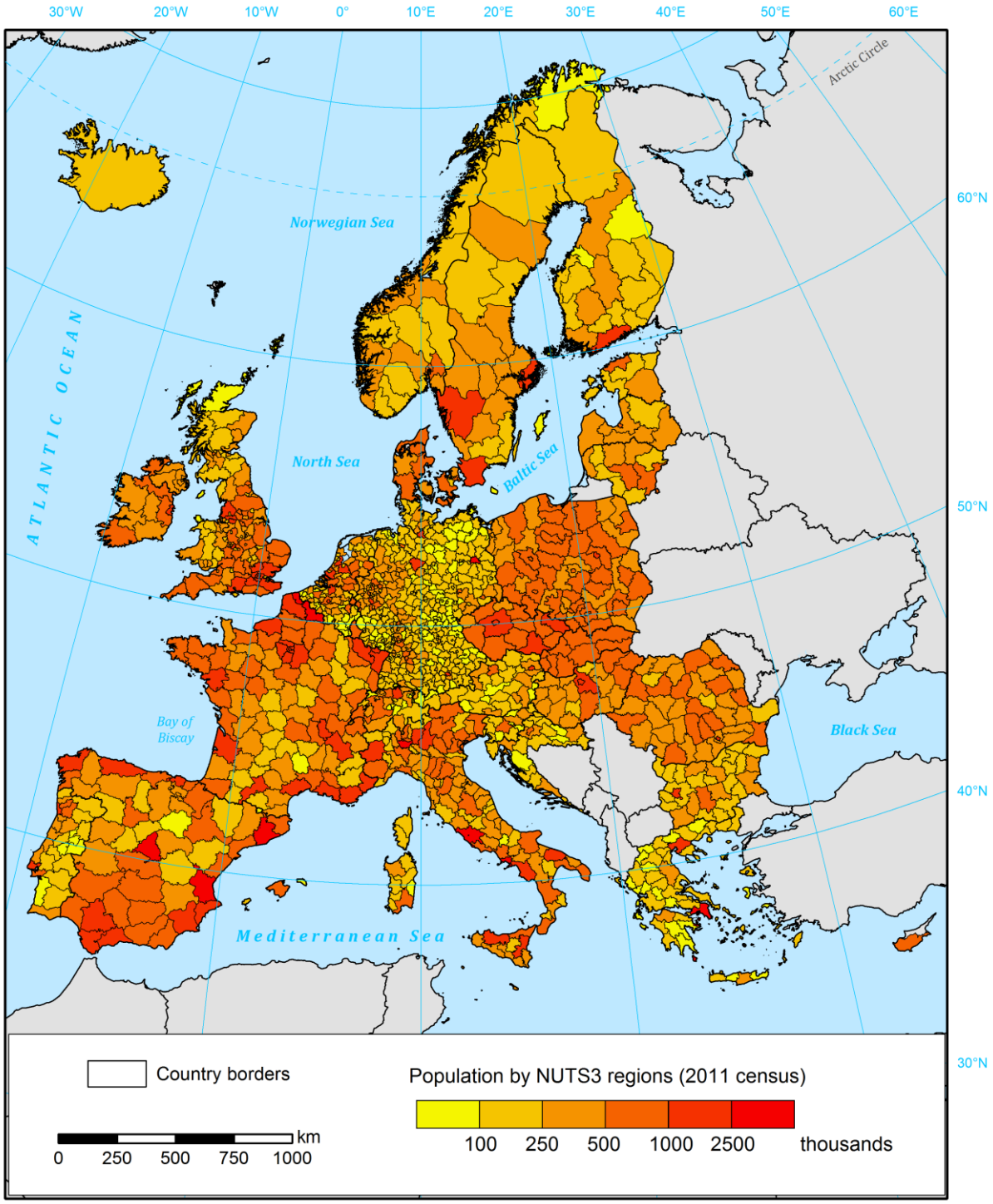

Figure 8. NUTS 3 regions (2010 version) in the study area and their population.

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)[3]. 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 1960-2010 data from Eurostat (2017) were mostly used, which included recalculation of historical census data to modern administrative divisions, and annual population estimates starting with 1990. Population projections up to 2020 were generally obtained from EUROPOP2013 projections by Eurostat (2017), except for countries with no subdivision into regions, which were obtained from newer EUROPOP2015 projections or from 2015 projections by United Nations (2015).

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*:

$$X_A^t = \sum F_{AB}\, X_B^t \tag{8}$$

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 order 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 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 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 3.2.2) 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

[3] 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.

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 3.2.1). 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 2000-2012, 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, are covered by croplands, pastures and forests was extrapolated using different data series following various definitions. For more recent years, regional data from Eurostat (2016) were largely used, otherwise national statistics, FAO (2016a) or HYDE 3.2 (Klein Goldewijk 2011) provided the necessary data. Statistics for forests were not collected for all years as they were needed only for validation, rather than land use modelling; additionally much less data is available compared to agricultural land. Area covered by road and rail infrastructure was extrapolated using statistics on motorway and railway length, mostly from national statistics, Eurostat (2017) and Mitchell (1981).

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 (Eurostat 2017). Here, we include estimates of GDP at constant prices, adjusted to 2011 price levels, with average currency exchange rates in 2011 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.

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. For years 2017-2020, the GDP data were extrapolated using latest (April 2017) projections by the International Monetary Fund (2017)..

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[4] (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 (2017), Goldsmith (1985) and several other compilations and country-specific sources. Historical series were linked to Eurostat's ESA 2010 estimates, where available. 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)[5];

[4] Potential inclusion of inventories, consumer durables and non-cultivated biological resources (mainly forests) was also reviewed. Those categories are destructible, and of considerable monetary value. However, very little data is available for those. An analysis of inventories and consumer durables data from OECD (2017), Goldsmith (1985), Piketty and Zucman (2014) and some other country-specific sources has shown that those assets are rather stable relative to GDP. Therefore, the omission of the assets shouldn't affect the analysis of trends in vulnerability to natural hazards.

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

- 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.

Value of dwellings and infrastructure 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).

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 7. Currency database format

Column	Description
Code	NUTS0 two-letter country code
Name	Country/territory name
Currency	Currency name*
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 purposes 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 2011 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 (2011 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 includes 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 four cases of hyperinflation: Germany 1923, Poland 1923, Greece 1944 and Hungary 1946. 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 three 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 mln pre-war Zlotys (PLO);
- Pre-war Zloty (PLO) was converted to post-war Zloty (PLL) at par (1:1) in 1944, then denominated to "heavy" Zloty (PLZ) at 100:3 in 1950, and again to "new" Zloty (PLN) in 1995 at 10,000:1. Additionally, the exchange rate to euro (EUR) in 2011 was 4.1206, hence

$$74600000 \,/\, \frac{1}{1} \,/\, \frac{100}{3} \,/\, \frac{10000}{1} \,/\, 4.1206 = 54.3125 \tag{9}$$

- Therefore, the uninflated value of losses equals 54.3125 EUR. From the GDP deflator series we can extract the price index for 1934, which is approx. 0.0000712, where year 2011 equals 100. Therefore:

$$54.3125 \times \frac{100}{0.0000712} = 76281571 \tag{10}$$

- Hence, the losses from the 1934 flood in 2011 prices can be estimated at 76.3 mln EUR.

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. Urban fabric and urban population redistribution (3.4.1);
2. Industrial or commercial units (3.4.2);
3. Reservoirs (3.4.12);
4. Infrastructure (3.4.3);
5. Airports (3.4.4);
6. Construction sites (3.4.5);
7. Croplands (3.4.8);
8. Pastures (3.4.9);
9. Burnt areas (3.4.10);
10. Natural areas (3.4.11);
11. Rural population redistribution (3.4.13);

The procedure is carried out separately in each NUTS 3 regions, and then the results are merged to create maps of land use and population.

3.4.1. Urban fabric (CLC 111 and 112) and urban population redistribution

Redistribution of population within urban areas and growth of cities was modelled based on two factors: change in urban population size and change in number of persons per households. Increasing population combined with smaller families in each dwelling have caused a substantial increase in demand for housing. Between 1870 and 2011, the number of urban households has increased 8-fold. Those extra dwelling had to be constructed 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). 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.

In light of the above, the modelling approach is as follows:

1. In every urban fabric grid cell *g* in region *r* the population *P* in time step *t* is modified relative to *t-1* (2011 baseline is step 0) to account for change is household size:

$$P_{t,r,g} = P_{t-1,r,g} \frac{H_{t,r}}{H_{t-1,r}} \tag{11}$$

where $H_{t,r}$ is the average number of persons per household, determined for each NUTS 3 region (see section 3.3.3);

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

$$S_{t,r} = \left(U_{t-1,r} - U_{t,r}\right)H_{t,r} - U_{t,r} \tag{12}$$

where $U_{t,r}$ is the urban population in the region according to the NUTS 3 database (see section 3.3.2);

4. If S_t is positive, it means that the urban area in time step *t* was smaller relative to *t-1*. Urban grid cells are removed starting with the lowest-ranked, and their population is removed as well, until the urban population in the region matches the desired value of $U_{t,r}$.
5. If S_t is negative, it means that the urban area in time step *t* was larger relative to *t-1*. Land use in non-urban grid cells are replaced by CLC 112 class starting with the highest-ranked. In each such grid cell, the population is increased to the threshold value of 65 persons (as defined in Table 3 in section 3.2.2), unless it is already higher than that. Urban areas are not allowed to sprawl into uninhabitable areas (as defined in Table 3).

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. Therefore, the distance from urban centre is a weighted sum of three Euclidean distances from:

- Centres of large agglomerations, as presented in a shapefile dataset from United Nations (2014), which shows the arbitrary centres of cities with a population larger than 300,000;
- Centroids of population clusters. Those clusters were calculated by Eurostat (2016b) from the 1 km population grid. The centroid was weighted, based on the population in each grid cell;
- Centroids of patches of urban fabric. The patches were taken from Corine Land Cover 2012, and centroids are based on the geometry of those patches.

Each of the three datasets was calculated separately for each region, using those "centres" which were located inside a rectangular envelope around each region (positioned at least 25 km from its borders). Each type of urban "centres" were given a different weight, based on a calibration process.

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

$$y = Ae^{-bx} \tag{13}$$

where *y* is the population density (in persons per km^2), *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 (for a complete overview, see Appendix 2). 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. A comparison of function parameters is presented in Fig. 10. Overall, a reasonable fit was achieved. It was found that an equal weighting of the three layers is most optimal. For cities for which more than one year of data was available, a decline of both parameters over time was observed, as in the literature case studies. A better match of modelled and observed estimates of eq. 10 parameters would be difficult, since the exponential curve fits are very sensitive to the sample size (not for all cities it was known within which distance population data were used) and the source material: literature studies used census wards of different sizes instead of a disaggregated population grid used here.

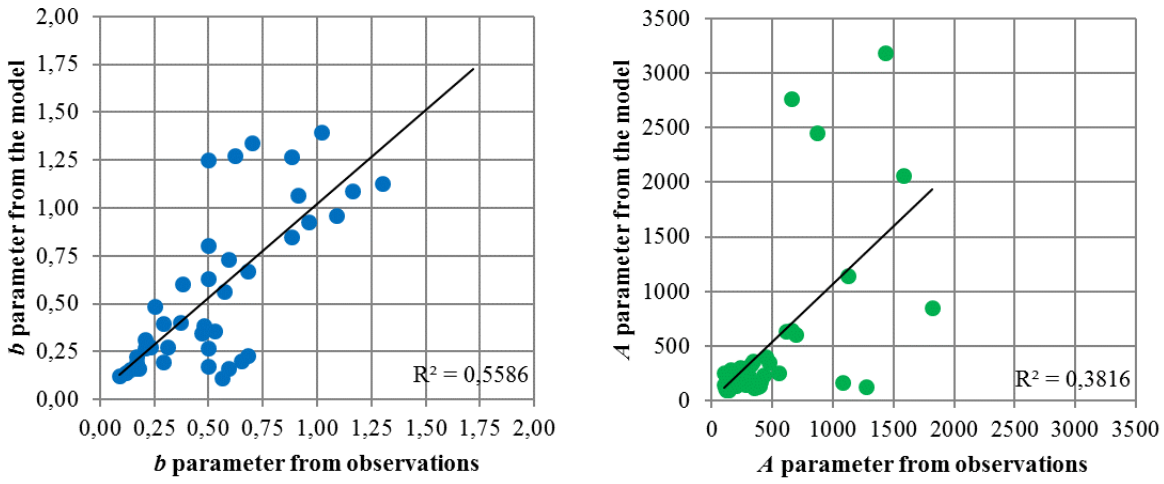

Figure 9. Estimates of* A *and* b *parameters (eq. 10) from modelled and observed population data.

3.4.2. Industrial or commercial units (CLC 121)

The area covered by large industrial facilities were assumed to change proportionately to industrial production per capita in constant prices (see section 3.3.7). The same approach as for urban fabric was applied (see previous section), with 'industrial' grid cells located furthest from the urban centres being removed first when going back in time.

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

The area covered by roads and railways was assumed to change proportionately to the length of motorways and railways (see section 3.3.4). The same approach as for urban fabric was applied (see section 3.4.1), with 'infrastructure' grid cells located furthest from the urban centres being removed first when going back in time.

3.4.4. Airports (CLC 124)

Airports first appeared in early 20th century. Due to the relatively small number of those objects in Europe (1,548) 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 (2016) open database, while the year of construction was gathered from various Internet resources. In some cases, the construction year was not available, therefore it had to be estimated based on available information, such as circumstances around their foundation or runway parameters[6]. Fig. 11 shows the number of airports identified, by year of construction.

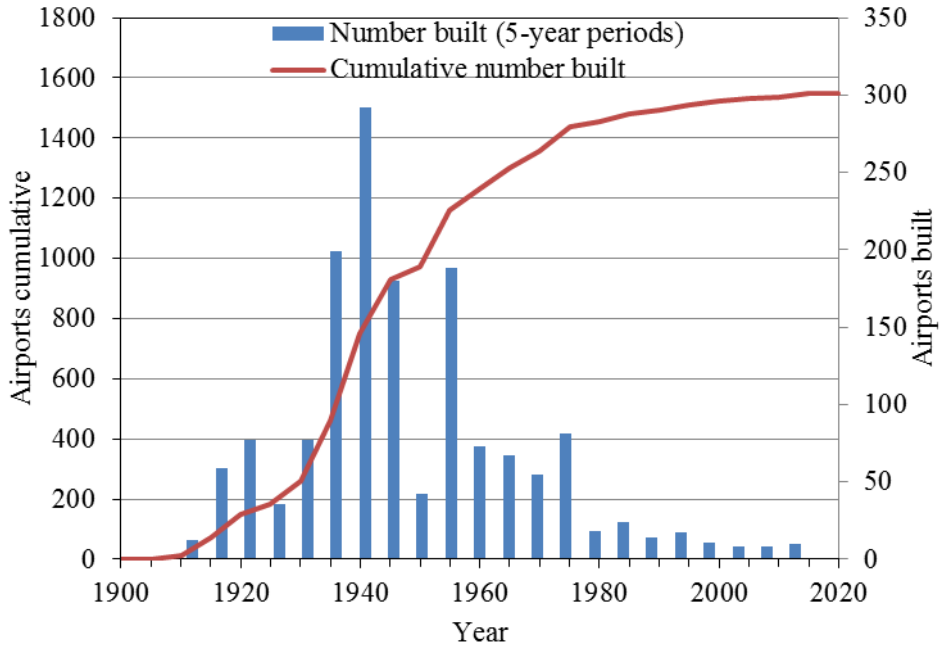

Figure 10. Number of airports built in the study area between 1900 and 2020.

3.4.5. Construction sites (CLC 133)

Construction sites are by definition a temporary land use, typically for only a few years. For years 2010–2020, their area was assumed constant, while for years 1870–2005 all construction sites were removed from the dataset.

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

Green urban areas, sport and leisure facilities are in principle uninhabited, and mostly unproductive and light in assets, therefore of little relevance for the analysis. In effect, their area was kept constant at every time step.

[6] For instance, substantial part of European airports were 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, 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.

3.4.7. 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 the of almost 15,000 objects. Therefore, their area was kept constant at every time step.

3.4.8. Croplands (CLC 211-223 and 241-244)

Modelling the changes in cropland area was based on an approach presented by Klein Goldewijk et al. (2011). It involves changing the allocation of croplands over time according to the land's suitability for agriculture. Therefore, if in time step *t* the cropland area was smaller than in time step *t-1*, 'cropland' grid cells are removed according to their ranking of suitability, starting with the lowest ranked cell (least favourable for croplands), until the value of cropland area in the NUTS 3 database is achieved (see section 3.3.4). Conversely, if in time step *t* the cropland area was larger than in time step *t-1*, 'non-cropland' grid cells are changed to CLC class 211 (non-irrigated agricultural land) starting with the highest ranked cell. The suitability is a sum of two indicators, which were also used by Klein Goldewijk et al. (2011)[7]. First is the slope of the terrain, which is serious limiter to agricultural activity, and which was calculated from EU-DEM dataset at 100 m resolution (Eurostat 2016b), see Appendix 3, Fig. A1). There is very close relationship between percentage of area used for croplands and slope, of exponential type (Fig. 12). The second indicator is the crop suitability index for high-input cereals as calculated by FAO (2016b) in the Global Agro-Ecological Zones (GAEZ) database (see Appendix 3, Fig. A2). The resolution of this dataset is 5' (about 4–7 km, depending on location). The index combines data on climate, soil and terrain to estimate potential yield of various crops. Out of several crops tested, high-input cereals have highest correlation with cropland fraction, of second-order polynomial type (Fig. 9).

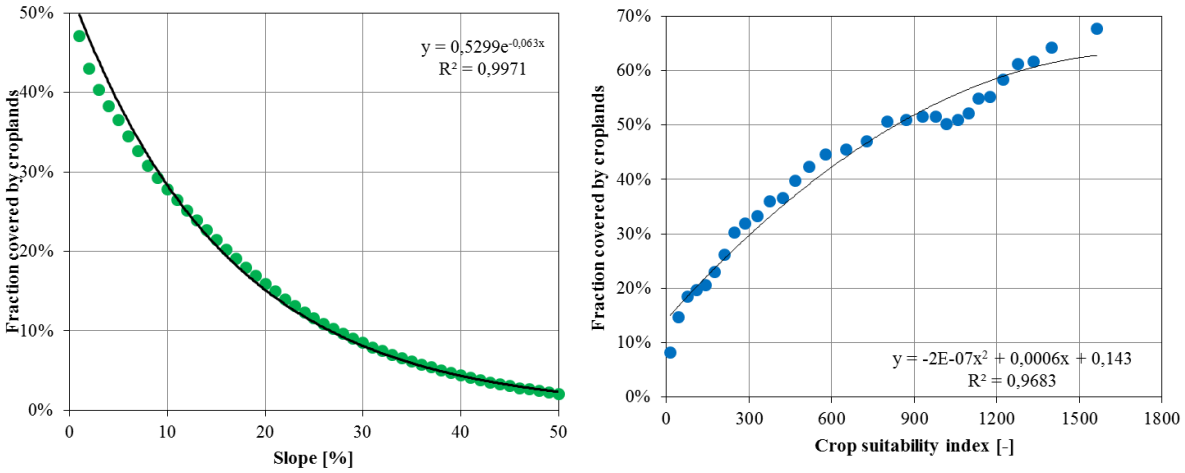

Figure 11. Fraction covered by croplands compared with slope (left) and crop suitability index for high-input cereals (right). Average fractions were calculated for slope divided into classes by rounding slopes to full percentages, and for crop suitability divided into 32 bins.

[7] The authors use more indicators, which were not applicable here due to different input data, extent and scope of this study.

For the slope indicator, the upper bound was set at 0% slope, while for the crop suitability index the upper bound was set at the polynomial function's maximum (approx. 1500). The suitability indicator for croplands I_c in a given grid cell is thus:

$$I_c = \frac{0.5299e^{-0.063S}}{0.5299} + \frac{-1.6 \cdot 10^{-7}C^2 + 5.6 \cdot 10^{-4}C + 0.143}{0.6327} \quad (14)$$

where *S* is the slope and *C* is the crop suitability index.

The main drawback of the method is that due to the relatively coarse resolution of the GAEZ dataset, there often many cells with the same rank, and the total area of croplands from the model does not exactly match the data in the NUTS3 database. Therefore, when too many cells have the same rank, they are further ranked by the centroid distance (see section 3.4.1), so that agricultural land with a given suitability class is added first closer to urban areas, and removed first furthest away from urban areas.

Additionally, the crop suitability index does not change over time, disregarding any variability of climate conditions, though Klein Goldewijk et al. (2011) considered this to be a valid assumption despite a much longer timespan of their study. The index refers to average conditions during years 1961–1990.

3.4.9. Pastures (CLC 231)

Modelling the changes in pastures follows the same methodology as croplands (see previous section). However, the crop suitability index for cereals was replaced by the same index for high-input alfalfa[8], a common crop growing on meadows and pastures (see Appendix 3, Fig. A3). The correlation is not as good as for croplands, but still has similar shape (Fig. 13).

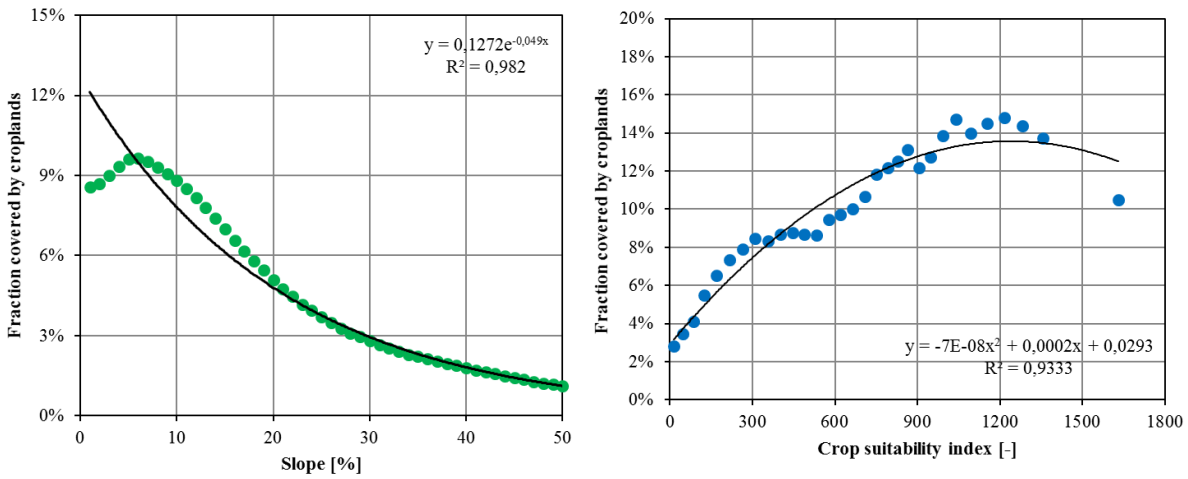

Figure 12. Fraction covered by pastures compared with slope (left) and crop suitability index for high-input alfalfa (right). Average fractions were calculated for slope divided into classes by rounding slopes to full percentages, and for crop suitability divided into 32 bins.

The suitability indicator for pastures I_p in a given grid cell is thus:

$$I_p = \frac{0.1272e^{-0.047S}}{0.1272} + \frac{-6.9 \cdot 10^{-8}C^2 + 1.7 \cdot 10^{-4}C + 0.0293}{0.1356} \quad (15)$$

[8] *Medicago sativa*, also known as lucerne.

where *S* is the slope and *C* is the crop suitability index. The index refers to average conditions during years 1961–1990.

3.4.10. Burnt areas (CLC 334)

Areas where vegetation has burned down (typically forests) are by definition a temporary land use. For years 2005–2020, burnt area was assumed constant, while for years 1870–2000 all burnt areas were removed from the dataset.

3.4.11. Natural areas, not covered by water (CLC 311–333 and 335–422)

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 land cover was defined as the most frequently occurring one within 200 m from the outline of the grid cell in question. If no natural land cover was located in the vicinity, the unoccupied land was assumed to be covered by forest (CLC 311), the most common natural land cover in Europe.

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

Areas covered by water are assumed to be constant over tide, thus not allowing for coastline and river course changes etc., expect for large reservoirs. A given reservoir was entirely removed from the land use dataset using information on the year of construction. 1069 reservoirs and their construction year were identified by intersecting Corine Land Cover 2012 with the Global Reservoir and Dam (GRanD) Database (Lehner et al. 2011). Fig. 14 shows the number of reservoirs identified, by year of construction.

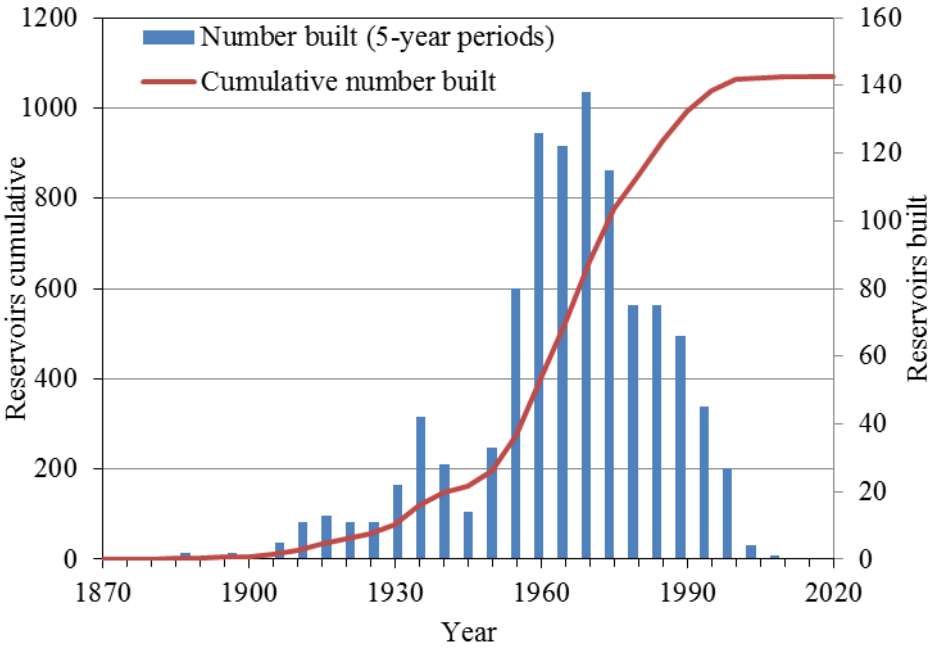

Figure 13. Number of large reservoirs built in the study area between 1870 and 2008. Based on data from Lehner et al. (2011).

3.4.13. Rural population redistribution

The final step of the procedure was to redistribute rural population in a procedure containing several steps, similarly to redistributing urban population. The steps were as follows:

1. For a given time step *t* and region *r*, the difference between rural population $R_{t,r}$ in non-urban grid cells (after application of all previous procedures in a given time step) and the rural population according to the NUTS 3 database $N_{t,r}$ was calculated:
$$W_{t,r} = R_{t,r} - N_{t,r} \quad (12)$$
2. If $W_{t,r} > 0$, the population of formerly urban grid cells *u*, which transitioned from urban to non-urban during the time step, was modified. Otherwise, this step was omitted. If the population of former urban grid cells was higher than the surplus, i.e. $\sum R_{t,r,u} > W_{t,r}$, the population number in all those cells was reduced by the same proportion, so that the rural population in the region would match the NUTS3 database:
$$R_{t,r,u} = R_{t,r,u} \frac{W_{t,r}}{\sum R_{t,r,u}} \quad (13)$$
3. If $W_{t,r} < 0$, the population number in all those cells was reduced to zero, i.e. $R_{t,r,u} = 0$.
4. Then, the population in all non-urban grid cells was modified according to the change in average household size, i.e.:
$$R_{t,r} = R_{t-1,r} \frac{H_{t,r}}{H_{t-1,r}} \quad (14)$$
where $R_{t,r}$ is the rural population in region *r* in time step *t*, and $H_{t,r}$ is the average household size.
5. In case there the realized $R_{t,r}$ and expected $N_{t,r}$ number of rural population are still different, population is added or subtracted iteratively, one person at a time to/from a inhabitable, non-urban grid cell (see section 3.4.1), starting with those closest to the urban centre, until $R_{t,r} = N_{t,r}$.

3.5. Disaggregation of economic data

Disaggregation of economic data is done in order to arrive to estimates of GDP and wealth per grid cell, just like the population and land use data. The methodology presented here extends the approach proposed by Milego and Ramon (2011) for the EU's ESPON 2013 Programme. In the ESPON study, and some others attempts such as G-Econ project (Nordhaus and Xi 2011), the GDP is disaggregated proportionally to the population. This approach works well with a relatively coarse resolutions of the output grid, however at 100 m resolution the economic variables are much less connected with the place of residence of the population. On the other hand, all economic activities require labour input. Therefore, using the observation that employee's compensation constitute approximately half of GDP in European countries (Eurostat 2017), GDP and wealth are disaggregated in equal proportion using population and land use.

Table 8 provides a summary of the assumptions behind the disaggregation. 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 the sectoral split at national level in 2011. The share of forestry in the agricultural sector varies from zero in Malta to 73% in Sweden.

Half of the GDP generated by agriculture (excluding forestry), as well as half of wealth (fixed assets) in this sector is distributed proportionally to the population living in agricultural areas. The other half was distributed equally among CLC classes 211-244 ("agricultural areas"). GDP and wealth in forestry was distributed the same way, but using CLC classes 311-313 ("forests"). Half of GDP and wealth in industry and services was distributed proportionally to the population in all grid cells, while the other half was distributed equally among specific land use classes where given production is concentrated, as in Table 8.

For the remaining two classes of wealth the approach was slightly different. The whole wealth in housing (dwellings) were distributed proportionally to the population in all grid cells. The entire value of infrastructure, on the other hand, was distributed equally over selected land use classes: urban fabric, airports, ports, roads and railway sites (CLC 111, 112, 122, 123 and 124).

Table 8. *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
GDP	Agriculture excl. forestry	Population in CLC211-244	CLC211-244
GDP	Forestry	Population in CLC311-313	CLC311-313
GDP	Industry	Total population	CLC121
GDP	Services	Total population	CLC111-121/133/141/142
Wealth	Housing	Total population	-
Wealth	Agriculture excl. forestry	Population in CLC211-244	CLC211-244
Wealth	Forestry	Population in CLC311-313	CLC311-313
Wealth	Industry	Total population	CLC121
Wealth	Services	Total population	CLC111-121/133/141/142
Wealth	Infrastructure	-	CLC111/112/122-124

3.6. Final exposure maps and analysis

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 2016. 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.

4.1. Criteria for inclusion for flood events

Flood events, in order to be included in HANZE-Events, had to fulfil certain criteria. The following rules were applied:

- At least one of the four statistics (area flooded, persons killed, persons affected, monetary value of losses) had to be available for a given event. However, if no persons were known to have been killed in the flood, at least one of the remaining statistics had to be available.

- Available information for a given event had to be good enough in order to assign month, year, country, regions affected, type and cause of the flood. Flood source (river/lake/sea name), detailed information on the cause or daily date were not required.
- Insignificant floods, which affected only a small part of one region, with no persons killed, were not included.
- Floods that were caused by insufficient drainage in urban areas not connected with any river system, floods caused entirely by dam failure unrelated with a severe meteorological event, or caused by geophysical phenomena were not included[9].

4.2. Database contents

HANZE-Events contains for each flood event relevant statistical information together with location, date, type, cause and other important information, as outlined in Table 9.

Table 9. Information contained in HANZE-Events database.

Variable	Description
No.	Event number
Country code	NUTS0 country code
Year	Year of the event (assigned from starting date)
Country name	Country in which the event occurred, using political divisions of the time of the event. In case of former countries of Czechoslovakia, East Germany, USSR and Yugoslavia, the appropriate succession state was used instead of the original country
Start date	Date on which the flood event started and ended; the exact daily dates are not always known, or are imprecise, but an event was included in the database as long as the starting month could be identified
End date	Date on which the flood event ended
Type	Type of flood event, which can be River, Coastal, River/Coastal, or Flash. The events were assigned to River/Coastal type if both factors contributed to the flooding. Flash flood type was assigned if the event was caused by rainfall lasting less than a day. However, often the information on meteorological conditions was missing and hence division of events into River and Flash floods was made based on dates of the event, location, season and impacts
Flood source	Name of the river, lake or sea from which the flood originated. The list of names is usually not comprehensive
Regions affected	Regions were flood damages were reported, using the NUTS3 delimitation of regions
Area flooded	Area inundated by the flood in km^2. This statistic more often than not relates only to agricultural land
Persons killed	Number of deaths due to the flood, including missing persons
Persons affected	Number of people whose houses were flooded. However, the reported numbers of persons affected often only show the number of evacuees or persons rendered homeless by the event. If no other number was available, those ones were used. If only the number of houses flooded was reported, the number persons affected was estimated by multiplying

[9] This exclusion pertains particularly to tsunamis, Icelandic *jökulhlaup* events (glacier melting by volcanic activity) or special events such as the 1963 Vajont Dam disaster.

	the number of houses by 4
Losses (nominal value)	Damages in monetary terms, in the currency and prices of the year of the flood event
Losses (mln EUR, 2011)	Damages in monetary terms converted to euro, correcting for price inflation relative to 2011
Cause	The meteorological causes of the event, including precipitation values, surge heights, etc. if available
Notes	Other relevant information, including co-occurrence of related events such as landslides or dam breaks, information on large discrepancies in the sources, estimated return periods and other relevant statistics
Sources	List of publications and databases from which the information was obtained

References

- 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
- 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
- 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) Pan-European, http://land.copernicus.eu/pan-european/. Last accessed 7 July 2017.
- 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.
- ESRI (2016) Europe Population Density, www.arcgis.com/home/item.html?id=cf3c8303e85748b5bc097cdbb5d39c31. Last accessed 20 December 2016.
- 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 (2016a) Administrative units / Statistical units, http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units. Last accessed 7 December 2016.
- Eurostat (2016b) GEOSTAT, http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat. Last accessed 8 December 2016.
- Eurostat (2017) Database, http://ec.europa.eu/eurostat/data/database. Last accessed 30 March 2017.

- FAO (2016a) FAOSTAT, http://faostat3.fao.org/download/R/RL/E
- FAO (2016b) Global Agro-Ecological Zones, http://gaez.fao.org/Main.html#
- 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
- Gisgraphy (2016) Download server, http://download.gisgraphy.com/openstreetmap/pbf/. Last accessed 8 December 2016.
- 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 Púčik, 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.
- International Monetary Fund (2017) World Economic Outlook Database, http://www.imf.org/external/pubs/ft/weo/2017/01/weodata/index.aspx. Last accessed 21 April 2017.
- 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 8 December 2016.
- 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
- 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
- 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 http://www.ums-riate.fr/Webriate/wp-content/uploads/2014/04/DB_TR_grids.pdf
- Mitchell, B. R. (1981) European historical statistics. Macmillan, London
- Nordhaus, W. and Xi, C. (2011) Geographically based Economic data (G-Econ), http://gecon.yale.edu/. Last accessed 3 April 2017.
- OECD (2017) OECD Data, https://data.oecd.org/
- OurAirports (2016) Open data downloads, http://ourairports.com/data/
- 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. (2017a) Efficient pan-European river flood hazard modelling through a combination of statistical and physical models. Natural Hazards and Earth System Sciences, in print, doi:10.5194/nhess-2017-4.
- Paprotny, D., Morales Nápoles, O., Vousdoukas, M. I., Jonkman, S. N., and Nikulin, G. (2017b) Accuracy of pan-European coastal flood mapping. Journal of Flood Risk Management, submitted.
- 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.
- 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 (2014) World Urbanization Prospects: The 2014 Revision. http://esa.un.org/unpd/wup/
- United Nations (2015) Demographic Yearbook 2014. United Nations, New York.
- 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. Detailed categories of non-financial assets

Table A1. 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	*Buildings other than dwellings*
AN.1122	*Other structures*
AN.1123	*Land improvements*
AN.113	Machinery and equipment
AN.1131	*Transport equipment*
AN.1132	*ICT equipment*
AN.1139	*Other machinery and equipment*
AN.114	Weapons systems
AN.115	Cultivated biological resources
AN.1151	*Animal resources yielding repeat products*
AN.1152	*Tree, crop and plant resources yielding repeat products*
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 2. Estimates of urban population density used in the analysis

Table A2. Estimates of urban population density. A, b – exponential function parameters (adjusted to give population density in persons per km2, 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	8	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	10	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	10	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	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
Dublin	IE021	1936	270	0,53	6	Clark 1951; Clark 1967
Frankfurt am Main	DE712	1890	550	1,16	5	Clark 1967
Frankfurt am Main	DE712	1933	340	0,57	7	Clark 1967
Leeds	UKE42	1951	116	0,31	10	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	5	Clark 1967
Stockholm	SE110	1940	425	0,48	8	Clark 1967
Vienna	AT130	1890	660	0,50	7	Clark 1951; Clark 1967
Zurich	CH040	1936	328	0,29	10	Clark 1967

Appendix 3. Supplementary maps

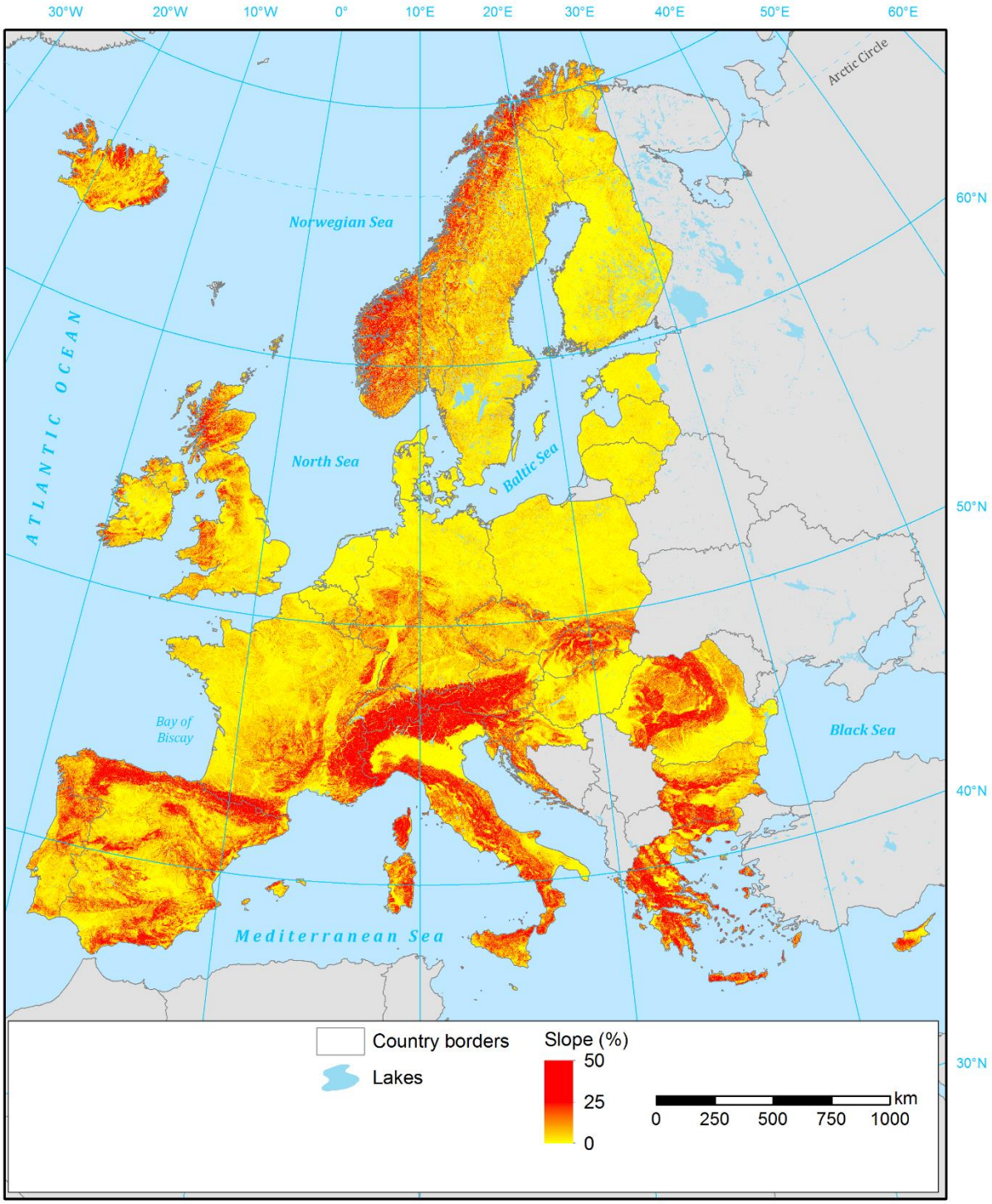

Figure A1. Slopeness of terrain in Europe, according to EU-DEM.

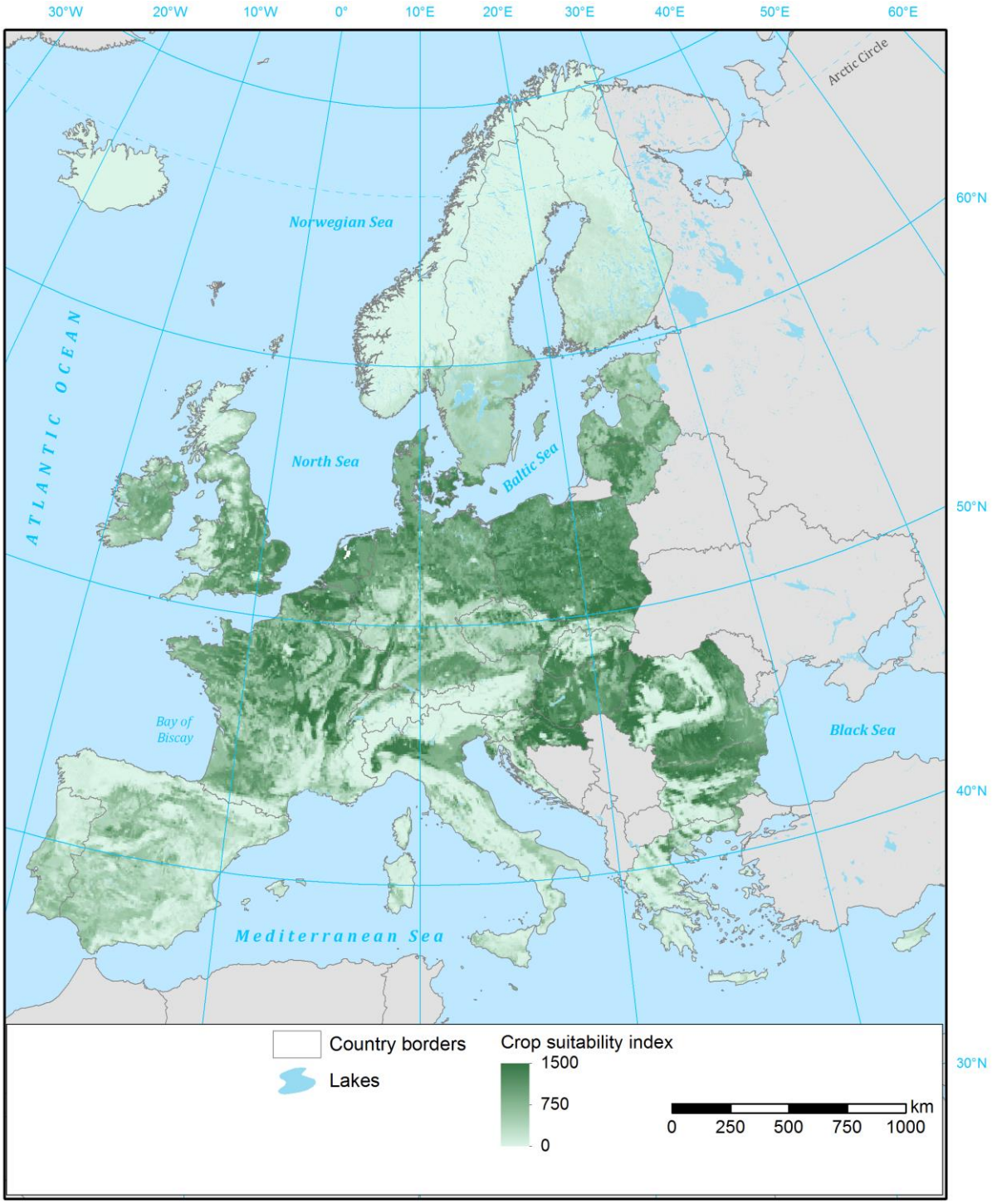

Figure A2. Crop suitability index for high-input cereals, according to GAEZ.

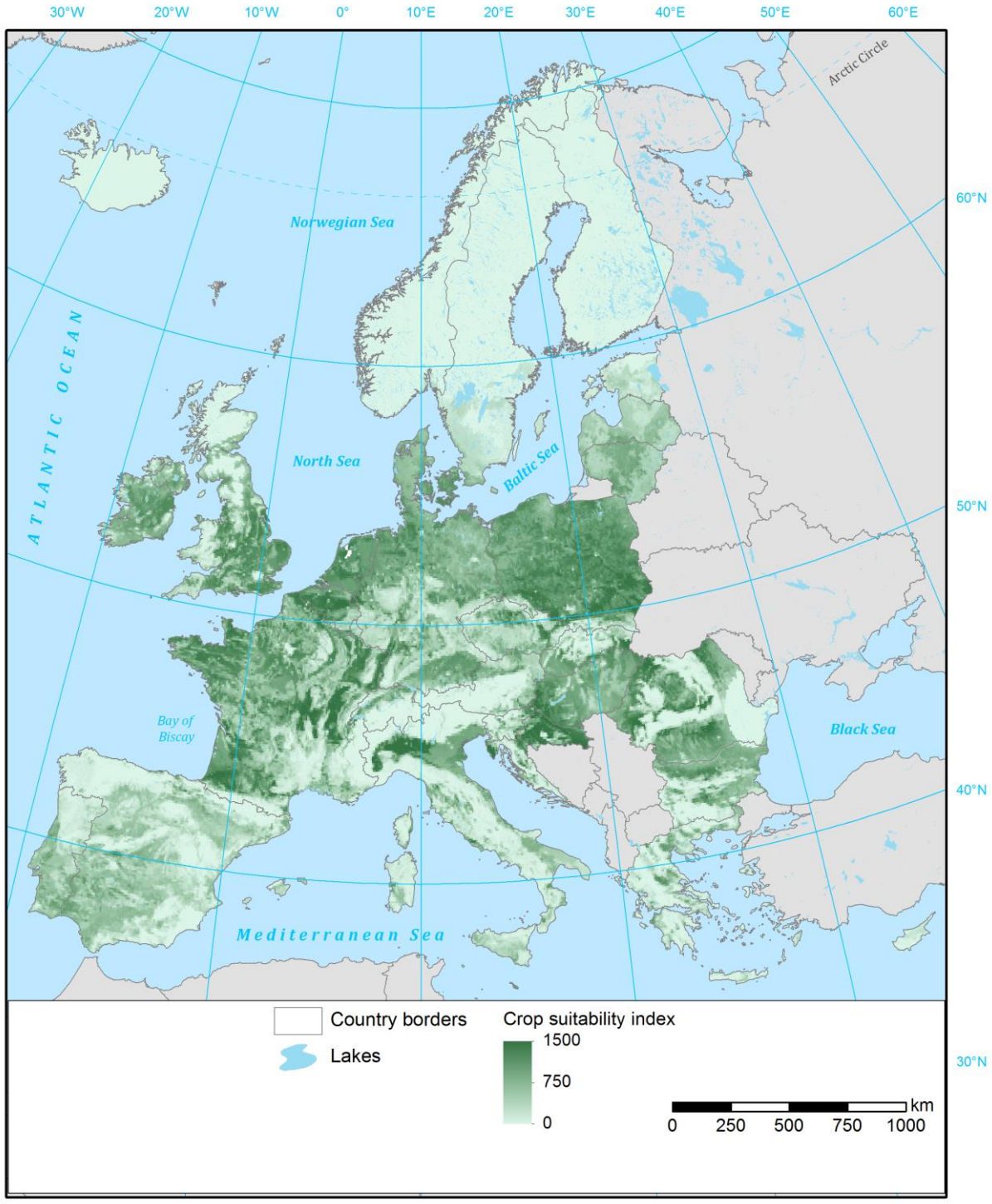

Figure A3. Crop suitability index for high-input alfalfa, according to GAEZ.