Open Access Earth System Science Data

HANZE: a pan-European database of exposure to natural hazards and damaging historical floods since 1870

Dominik Paprotny, Oswaldo Morales-Nápoles, and 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: Dominik Paprotny (d.paprotny@tudelft.nl)

Abstract. The influence of social and economic change on the consequences of natural hazards has been a matter of much interest recently. However, there is a lack of comprehensive, high-resolution data on historical changes in land use, population, or assets available to study this topic. Here, we present the Historical Analysis of Natural Hazards in Europe (HANZE) database, which contains two parts: (1) HANZE-Exposure with maps for 37 countries and territories from 1870 to 2020 in 100 m resolution and (2) HANZE-Events, a compilation of past disasters with information on dates, locations, and losses, currently limited to floods only. The database was constructed using high-resolution maps of present land use and population, a large compilation of historical statistics, and relatively simple disaggregation techniques and rule-based land use reallocation schemes. Data encompassed in HANZE allow one to "normalize" information on losses due to natural hazards by taking into account inflation as well as changes in population, production, and wealth. This database of past events currently contains 1564 records (1870–2016) of flash, river, coastal, and compound floods. The HANZE database is freely available at https://data.4tu.nl/repository/collection:HANZE.

1 Introduction

Natural hazards take place when recurring extremes of the Earth's environment collide with human activities. Beyond the natural or anthropogenic changes to the environment, the extent of those activities has profound effects on the consequences of disasters. Even in a span of a few decades, social, economic, and technological developments drive the constant evolution of exposure and vulnerability to hazards. Therefore, there is growing interest in how much the number of persons and assets at risk has changed over time worldwide (Jongman et al., 2012; Kummu et al., 2016; Schumacher and Strobl, 2011), and what consequences those findings have for observed trends in natural hazards-related losses (Bouwer, 2011; Bouwer et al., 2007; Daniell et al., 2011; Munich Re, 2016; Schiermeier, 2006).

Floods in Europe have received particular attention (Barredo, 2007). Barredo (2009) found that correcting reported flood losses for inflation and economic growth yields no trend for 1970–2006, in contrast to steep rise in orig-

inally reported losses. Similar findings were presented for the United Kingdom, covering years 1884–2013 (Stevens et al., 2016). Other studies on trends in flood exposure were carried out, e.g. for Austria (Fuchs et al., 2015b), Italy (Domeneghetti et al., 2015), the Netherlands (Jongman et al., 2014), Spain (Barredo et al., 2012), Switzerland (Röthlisberger et al., 2016), and the United Kingdom (Stevens et al., 2015). The importance of population and economic growth, but also land use distribution, has been emphasised (Boudou et al., 2016; Sofia et al., 2017). At the same time, information on past flood losses are being collected in national (Guzzetti and Tonelli, 2004; Haigh et al., 2015) and international databases (Brakenridge, 2017; Guha-Sapir et al., 2017; Munich Re, 2017), including data collected as part of European Union-mandated preliminary flood risk assessments (European Environment Agency, 2015).

However, there are several limitations of the aforementioned studies and databases. Exposure data sets were derived at a variety of spatial and temporal resolutions with differ-
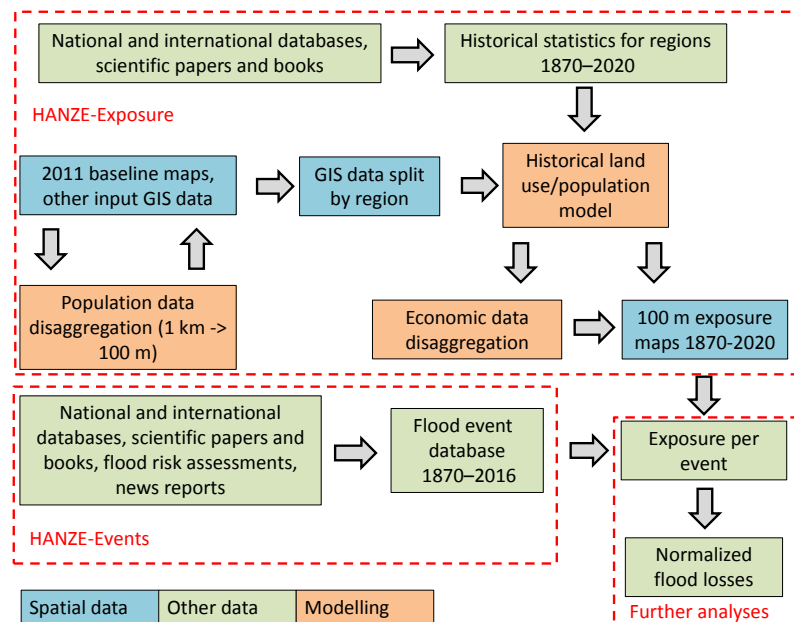
Figure 1. Workflow of the HANZE database from input data sets to final exposure maps and flood events database, and an example of how the two components interact to derive normalized flood losses.

ent thematic coverage. Within a given country, typically one series of population, gross domestic product (GDP), housing stock, or other variable were used to normalize reported flood losses. This approach neglects substantial variation in development within countries. Also, the availability of past flood damage information is very uneven between countries, and international databases only provide reasonable coverage beginning in the 1980s. The timespan of the studies on exposure is usually limited to the most recent decades, given the lack of adequate data.

A typical source of gridded historical population and land use is HYDE (Klein Goldewijk et al., 2017), which has a 5′ resolution (approx. 40–60 km^2 over Europe) and a very long time span, from 10 000 BC to AD 2100. HYDE utilizes historical population estimates combined with a set weighting maps for land use to generate gridded reconstructions of the past anthropogenic environment. Other time-varying data sets of gridded population include GPW v4 for years 2000–2020 (CIESIN/SEDAC, 2018) and GHSL for years 1975–2014 (Pesaresi et al., 2016). Disaggregated GDP is provided for years 1990–2005 by GEcon 4.0 (Nordhaus and Xi, 2011) and for years 1980–2100 using the Murakami and Yamagata (2017) data set. Finally, relatively detailed, 1 km resolution maps of European land cover/use were created for 1900 (Fuchs et al., 2015a) and 1950–2010 (Fuchs et al., 2013). Still, there is a lack of comprehensive data sets that would allow for normalizing losses from past natural hazards, especially those needed to be analysed at very fine resolution, like floods.

Drawing from recent developments in pan-European demographic and land use mapping, as well as new studies on historical changes in population, production, and wealth, we seek to address the aforementioned weaknesses with a new comprehensive data set. Historical Analysis of Natural Hazards in Europe (HANZE) 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, namely HANZE-Exposure and HANZE-Events. HANZE-Exposure consists of high-resolution gridded data with information on land use, population, production, and wealth per 100 m grid cell from 1870 to 2020. It allows one to derive potential damages for any past natural hazard with a defined spatial extent. The other component, HANZE-Events, contains information on location, time, and quantitative data on consequences of past natural disasters, currently limited to floods (1870–2016). It is supplemented by economic data necessary for converting nominal monetary losses into a single benchmark. HANZE covers 37 European countries and territories constituting approximately 70 % of the continent's population (Eurostat, 2017). The composition of the domain is detailed in Supplementary File 1 and Fig. S1 in this file.

As presented in Fig. 1, the starting points for constructing HANZE-Exposure database were a gridded land cover/use map (100 m resolution) and a population map (1 km resolution), both covering the situation in Europe ca. 2011. Based on previously published methods, demographic and economic data were disaggregated to 100 m resolution, and changes in historical land use and population were modelled

utilizing a large compilation of historical statistics at the regional level. HANZE-Events was created from a wide array of published sources and databases. The end date of HANZE-Exposure is different from HANZE-Events, because exposure data are prepared with a 10-year time step for 1870–1970 and a 5-year time step for 1970–2020. Therefore, a short-term projection for 2020 is necessary to calculate exposure for post-2015 events. It should be noted that the starting year of 1870 was chosen mainly due to data availability.

2 Methods

The creation of HANZE-Exposure data involved four major steps, which are explained below. Main sources and concepts for HANZE-Events are outlined afterwards.

2.1 Exposure step 1: baseline maps

There are very few high-resolution population and land cover/land use maps, and data sets constructed with a certain methodology rarely extend beyond a single time point. Therefore, two maps (one each for population and land cover) for a single year (2011 or 2012) were collected as baseline for the study. All other time points between 1870 and 2020 are calculated from those baseline maps using historical statistics with substantially lower resolution.

The baseline land cover/use is based on CORINE Land Cover (CLC) 2012, version 18.5a (Copernicus Land Monitoring Service, 2017). CLC is a project supervised by the European Environment Agency. It has so far produced four pan-European land use maps for 1990, 2000, 2006 and 2012. The maps are prepared mostly by manual classification of land cover patches from satellite imagery with a resolution of 25 m or better. For the latest edition, images collected during 2011–2012 were used. The inventory consists of 44 classes (Fig. S3). The minimum size of areal features is 25 ha. For linear objects such as roads, railways, and rivers, a minimum width of 100 m is used. CLC 2012 is first displayed as a vector map, and can then be transformed into a raster with 100 m resolution. CLC 2012 covers the entire domain with the exception of Andorra. For this particular country, the land cover/use map was constructed with overlaying data from four different sources, top to bottom:

1. CLC 2012 v18.5a, which covers a small strip around the border;

2. CLC 2000 v18.5, an earlier edition which covers a larger strip around the border (Copernicus Land Monitoring Service, 2017);

3. OpenStreetMap, accurate as of mid-2016 (Gisgraphy, 2017);

4. Global Land Cover 2000 (Joint Research Centre, 2015).

The final map for the full domain of 37 countries and territories is presented in Fig. S2.

The baseline population map is based on the GEOSTAT 2011 population grid, version 2.0.1 (Eurostat, 2017). This data set has 1 km resolution and for most countries it represents the actual population enumerated and georeferenced during the 2011 round of population censuses, complemented by estimates by the European Commission's Joint Research Centre. This data set is presented in Fig. S4. For this study, the 1 km grid had to be further disaggregated to 100 m resolution. Several methods have been proposed for this procedure and tested for Europe (Gallego, 2010; Gallego et al., 2011). 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. The procedure is an iterative algorithm applied separately for each 1 km grid cell. The steps are as follows:

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

$$Y_{LG}^0 = Y_G = \frac{X_G}{S_G}, \tag{1}$$

where Y_{LG}^0 is the population density for land use $L \in \{1, \ldots, n\}$ 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.

2. A population density threshold T_L is defined for each one of n land use classes.

3. 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.

4. 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:

$$U_{LG}^L = S_{LG} \cdot \left(Y_{LG}^{L-1} - T_L \right). \tag{2}$$

5. Surplus is then redistributed among the remaining land use classes M; hence

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

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

6. 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:

$$Y_{LG} = \frac{T_L X_G}{\sum T_L S_{LG}}. \tag{5}$$

The crucial aspect of this method is defining the threshold T_L. Here, we use thresholds 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. The final thresholds T_L are shown in Table S1 in the Supplement. 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. It should be noted that land use classes with $T_L = 0$ are still included in the algorithm described above.

The result of the calculation, however, is only the population per land use in each 1 km grid cell. Hence, the population had to be disaggregated further. For this 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, to 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 data set (Copernicus Land Monitoring Service, 2017). It was created based on high-resolution satellite photos taken during 2011–2012 in visible and infrared spectrum. This data set has 100 m 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 law function, based on cell imperviousness ranging from 1 to 96 % (Fig. S5). Cells with 0 % imperviousness should, in principle, not be inhabited. Additionally, a power law function converges at 0 %. At the opposite on scale, almost no 1 km cells have values above 96 %. Hence, the population X_g in 100 m grid cell g is equal to

$$X_g = \left[\frac{Z_g}{\sum Z_g} Y_{LG} S_{LG} \right], \tag{6}$$

where Z_g is the population of grid cell g obtained from the power function divided by the 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 to the closest integer, 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 increased or reduced randomly (with equal probability) within the land use class, one person at a time, until the population X_{LG} matches the value before the second stage of disaggregation. This completes the process, an example of which is shown in Fig. 2.

2.2 Exposure step 2: historical statistics

Reconstruction of exposure for years other than the baseline maps requires historical statistics for several variables. Most of those statistics have been collected at regional level. The Nomenclature of Territorial Units for Statistics (NUTS), 2010 edition (European Union, 2011), was used here to define the region. This classification has four levels (0, 1, 2, 3), where 0 is the national level and 3 is the finest regional division. Level 3 was chosen for this study, resulting in 1353 regions in the study area (Fig. S6). A vector map of regions was obtained from ESRI (2016) with amendments based on Eurostat (2017) map in order to fully match NUTS 2010 classification. Coastlines in the vector map were further adjusted using the aforementioned CLC 2012 map. NUTS favours administrative divisions in defining the regions, though often statistical (analytical) regions are used instead, created by amalgamation of smaller administrative units. It should be noted that NUTS 2010 was used instead of newer editions because 2011 census data, matching the baseline population map, were disseminated using this classification of regions.

All variables collected and used as input to HANZE-Exposure are listed in Table 1. Detailed definitions and concepts for all variables are include in Supplementary File 1. Their utility for the study is explained in the subsequent subsections. In general, all variables were collected from almost 300 sources, so that a time series for one variable for one country was typically merged from several sources. Due to the number of sources and transformations required to complete the database, only the most important methods and sources are mentioned in the Supplementary File 1. Full descriptions of sources and methods are included per country, separately for each variable, with the exception of the "forestry index", "airports", and "reservoirs" variables, which are described in this manuscript as they were compiled in a more straightforward manner.
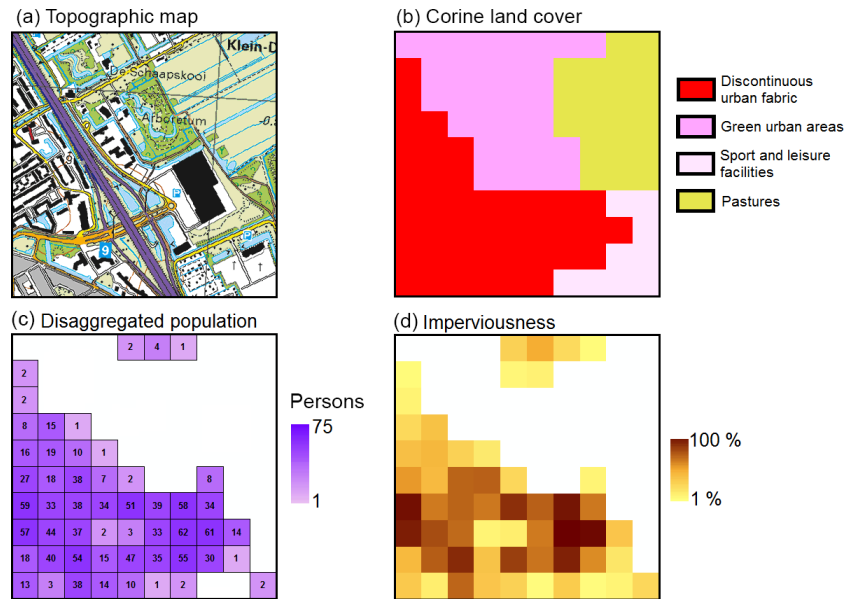
Figure 2. Disaggregation result and source data for a fragment of the city of Delft in the Netherlands. The area shown corresponds to a 1 km grid in the GEOSTAT population data set. In this grid cell, the population at the time of the 2011 census was 1218. Panel (**a**) was extracted from a Dutch 1 : 25 000 topographic map for reference. Panel (**b**) shows the land use structure according to CORINE Land Cover 2012, and panel (**d**) shows soil sealing according to Imperviousness 2012 data set. The final disaggregated 100 m population grid is presented in panel (**c**).

Table 1. Input variables in HANZE-Exposure.

Variable	Unit	Resolution
Total population	thousands of persons	regions, 5-/10-year
Urban fraction	urban population as a percent of total population	regions, 5-/10-year
Persons per household	mean number of persons	regions, 5-/10-year
Croplands	percent of total area	regions, 5-/10-year
Pastures	percent of total area	regions, 5-/10-year
Infrastructure	area covered by road and rail infrastructure in ha	regions, 5-/10-year
GDP	millions of euros in constant 2011 prices	regions, 5-/10-year
GDP from agriculture	percent of total GDP	regions, 5-/10-year
GDP from industry	percent of total GDP	regions, 5-/10-year
GDP from services	percent of total GDP	regions, 5-/10-year
Wealth in housing	percent of total GDP	countries, 5-/10-year
Wealth in agriculture	percent of GDP from agriculture	countries, 5-/10-year
Wealth in industry	percent of GDP from industry	countries, 5-/10-year
Wealth in services	percent of GDP from services	countries, 5-/10-year
Wealth in infrastructure	percent of total GDP	countries, 5-/10-year
Forestry index	percent of GDP from agriculture	countries, 2011
Airports	year of construction	CLC patches, annual
Reservoirs	year of construction	CLC patches, annual

2.3 Exposure step 3: land use and population change modelling

After the baseline maps and a database of historical statistics were completed, changes in land use and population over time were modelled. This was carried out for each of the 1353 NUTS 3 regions separately in specified order. A summary of the procedure is included in Table 2 and the most important details of the methodology are described below.

Redistribution of population within urban areas and growth of cities were 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

Table 2. Summary of historical land use and population modelling approaches, by CORINE Land Cover classes (see Fig. S3). The number in first column indicates the order in which the modification of land use and population was done.

Order	Land use and population type	Modelling approach
1	urban fabric (CLC 111 and 112) and urban population redistribution	Population per urban grid cell is modified according to changes in mean number of persons per household. Surplus population (the difference between urban population in a region after this modification and the value reported in the historical statistics database) and urban fabric are removed starting with grid cells furthest away from urban centers (see text for details).
2	industrial or commercial units (CLC 121)	Area of CLC 121 in a region changes proportionately to industrial production per capita in constant prices. "Industrial" grid cells located furthest from the urban centres are removed first when going back in time.
3	reservoirs (part of CLC 512)	Reservoirs are removed completely using the information on year of construction. 1069 objects and their construction year were identified using GRanD database (Lehner et al., 2011).
4	road and rail networks and associated land (CLC 122)	Area of CLC 122 in a region changes as defined in the historical statistics database. "Infrastructure" grid cells located furthest from the urban centres are removed first when going back in time.
5	airports (CLC 124)	Airports are removed completely using the information on year of construction. 1548 objects were identified using mostly OurAirports (2017) database and year of construction was mostly obtained from various language editions of Wikipedia.
6	construction sites (CLC 133)	All construction removed from the land use map for years 1870–2005, otherwise as in the baseline map.
7	croplands (CLC 211–223 and 241–244)	The area covered by croplands in a region is adjusted to match the value in the historical statistics, so that the grid cells least suitable for agriculture are removed first, while unutilized grid cells with the highest suitability are added first. Suitability is proportional to slope and crop suitability index for high-input cereals by FAO (2016). Grid cells ranked the same are disambiguated with distance from urban centres (see text for details).
8	pastures (CLC 231)	As for croplands, but with crop suitability index for high-input alfalfa used instead of cereals (see text for details).
9	burnt areas (CLC 334)	All burnt areas removed from the land use map for years 1870–2000, otherwise as in the baseline map.
10	natural areas other than water (CLC 311–333 and 335–422)	If after application of previous steps some land becomes unoccupied, it is assumed that this land was covered by the same natural land cover typical to its nearest neighbourhood (the most frequently occurring type 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).
11	rural population redistribution	Population of grid cells which were changed from urban to non-urban is modified, then non-urban population is modified according changes in mean number of persons per household. If needed, rural population increased/reduced based on distance from urban centres to match historical statistics for a region (see text for details).
12	remaining land use (CLC 122, 131, 132, 141, 142, 423, 511, and 521–523)	Assumed constant, as in the baseline map.

in the demand for housing. Between 1870 and 2011, the number of urban households in Europe increased eight-fold. Those extra dwellings were typically constructed outside the urban centres, as existing houses were rarely replaced by bigger ones. Many studies have shown a functional relationship between population density and distance from the city centre (Berry et al., 1963; Anas et al., 1998; Papageorgiou, 2014). Clark (1967) showed 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 existing households, families have became smaller, and thus the population declines closer to the centre and the surplus population is accommodated further from the centre in less-developed areas.

In light of the above, the modelling procedure 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 in household size:

$$P_{t,r,g} = P_{t-1,r,g} \frac{H_{t,r}}{H_{t-1,r}}, \tag{8}$$

   where $H_{t,r}$ is the average number of persons per household in each region.

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 as

$$S_{t,r} = \left(U_{t-1,r} - U_{t,r}\right) H_{t,r} - U_{t,r}, \tag{9}$$

   where $U_{t,r}$ is the urban population in the region according to the historical statistics database.

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 S1), unless it is already higher than that. Urban areas are not allowed to sprawl into uninhabitable areas (Table S1).

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 distributions far outside their borders. Therefore, the distance from urban centre is a weighted sum of three Euclidean distances from the following:

– Centres of large agglomerations, as presented in a shapefile data set from United Nations (2014), which shows the arbitrary centres of cities with a population larger than 300 000.

– Centroids of population clusters. These clusters were calculated by Eurostat (2017) from the 1 km population grid. The centroids were weighted, based on the population in each grid cell.

– Centroids of patches of urban fabric. The patches were taken from CORINE Land Cover 2012 (Copernicus Land Monitoring Service, 2017), and centroids are based on the geometry of those patches.

Equal weighting of the three layers was found to be optimal by analysing the approach's accuracy (see Sect. 3.2). After urban fabric and population are redistributed, changes in area covered by other types of artificial surfaces, as well as reservoirs, are accounted for (see Table 2). Then, evolution in cropland area is modelled using an approach similar to one utilized in HYDE database of historical land use and population (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 suitable for croplands), until the value of cropland area in the historical statistics database is achieved. 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 in the HYDE database. The first indicator is the slope of the terrain (Fig. S7), which is a serious limiter on agricultural activity, and which was calculated from EU-DEM data set at 100 m resolution (Eurostat, 2017). We found a close exponential-type relationship between percentage of area used for croplands and slope. The second indicator is the crop suitability index for high-input cereals as calculated by FAO in the Global Agro-Ecological Zones (GAEZ) database (FAO, 2016; Fischer et al., 2002). The resolution of this data set is $5'$ (about 40–60 km^2, depending on location). The index combines data on climate (1961–1990), soil, and terrain to estimate potential yield of various crops. Out of several crops tested, high-input cereals (Fig. S8) have highest (second-order polynomial) correlation with cropland fraction.

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

$$I_c = \frac{0.5299e^{-0.063S}}{0.5299}$$
$$+ \frac{-1.6 \times 10^{-7}C^2 + 5.6 \times 10^{-4}C + 0.143}{0.6327}, \tag{10}$$

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 data set, there are often many cells with the same rank, and the total area of croplands from the model does not exactly match the data in the historical statistics database. Therefore, when too many cells have the same rank, they are further ranked by the centroid distance (as for urban population), so that agricultural land with a given suitability class is the first added closest to urban areas, and is the first removed furthest away from urban areas.

Modelling the changes in pastures follows the same methodology as croplands, except that the crop suitability index for cereals was replaced by the same index for high-input alfalfa (also known as lucerne), a common crop growing in meadows and pastures (Fig. S9). The suitability indicator for pastures I_p in a given grid cell is thus

$$I_p = \frac{0.1272e^{-0.047S}}{0.1272} + \frac{-6.9 \times 10^{-8}C^2 + 1.7 \times 10^{-4}C + 0.0293}{0.1356}. \quad (11)$$

The indicators and functional relationships used for analysing agricultural land use changes are included in Figs. S10 and S11. After modelling croplands and pastures, burnt areas are removed where necessary (see Table 2) and unoccupied land is replaced by natural vegetation. The final step is thus the redistribution of rural population. The procedure, similar to one employed for urban population, is 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 as

$$W_{t,r} = R_{t,r} - N_{t,r}. \quad (12)$$

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 the case that the realized $R_{t,r}$ and expected $N_{t,r}$ numbers of rural population are still different, population is increased or reduced iteratively, one person at a time to/from a inhabitable, non-urban grid cell (CLC classes 211 to 324; see Table S1), starting with those closest to the urban centre, until $R_{t,r} = N_{t,r}$.

2.4 Exposure step 4: disaggregated economic data

Disaggregation of economic data provides estimates of GDP and wealth per grid cell, just like the population and land use data. It was carried out after historical gridded population and land use were obtained. The methodology presented here extends the approach proposed in the European Union's ESPON 2013 Programme (Milego and Ramos, 2011) and some others studies, such as G-Econ project (Nordhaus and Xi, 2011), in which the GDP is disaggregated proportionally to the population. This approach works well with a relatively coarse resolution 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 still require labour input. Using the observation that employee's compensation constitutes approximately half of GDP in European countries (Eurostat, 2017), the GDP and wealth are disaggregated in equal proportion using population and land use. It should be noted that wealth is defined here as tangible, produced, non-financial fixed assets. The composition of wealth is detailed in the Supplement and Table S2.

Table 3 provides a summary of the assumptions behind the disaggregation. Additional assumptions had to be made for the agricultural sector, which is the most dispersed, as almost three-quarters of the study area are 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, a 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. Hence, agricultural GDP and wealth at the regional level were broken down to forestry (including logging) and remaining agriculture (including fishing and aquaculture) using the sectoral split at national level in 2011 from Eurostat (2017). The share of forestry in the agricultural sector varies from zero in Malta to 73 % in Sweden.

Table 3. Disaggregation of economic variables by population and land use classes (CLC: CORINE Land Cover).

Variable	Category	Disaggregated into population	Disaggregated into 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

Half of the GDP generated by agriculture (excluding forestry), as well as half of the wealth 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 by using CLC classes 311–313 ("forests"). Half of the GDP and half of the wealth in industry and services were distributed proportionally to the population in all grid cells, while the other halves were distributed equally among specific land use classes where a given production is concentrated, as in Table 3.

For the remaining two classes of wealth, the approach was slightly different. The entire wealth in housing (dwellings) was 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 and 122–124).

2.5 Database of flood events

HANZE-Events includes information on past damaging floods that occurred in the domain (37 countries and territories) between 1870 and 2016. Several rules were applied to determine whether a flood event indicated in sources should be included in the database, as follows:

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

– Insignificant floods, i.e. events which affected only a small part of one region, with no fatalities and less than 200 persons affected, were not included.

– Available information for a given event had to be sufficient in order to assign month, year, country, regions affected, type of flood, and general cause of the event. Flood source (river/lake/sea name), detailed information on the cause and day of the event were not required.

– 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 (such as tsunamis or *jökulhlaup* events) were not included.

– Flood events that had impact on more than one country were split per country as long as data were available on per country basis. Otherwise they were presented as one flood event. Also, in the case of an event affecting several regions of a country, when the availability of statistics per region is uneven, the event was split accordingly.

Records of flood events were obtained from a large variety of sources (more than 300), including international and national databases, scientific publications, and news reports. The source of information is indicated per event in the HANZE-Events data set. In the majority of cases, entries taken from international databases were cross-checked with other sources and amended as necessary. Databases particularly worth mentioning are EM-DAT (Guha-Sapir, 2017), Dartmouth Flood Observatory (Brakenridge, 2017), NatCatService (Munich Re, 2017), European Environment Agency database of historical information submitted under Floods Directive (2015), the national flood databases of France (Lang et al., 2016), Italy (Guzzetti and Tonelli, 2004), Spain (Dirección General de Protección Civil, 2015), and the United Kingdom (Black and Law, 2004; Haigh et al., 2015), and several national and regional preliminary flood risk assessments.

In order to convert reported losses from various currencies and reference years to a single benchmark, information on inflation and currencies were collected. Two tables were prepared and are included with other HANZE input data. The first one includes all currencies that were used in the study area between 1870 and present, with their names, ISO 4217 codes, starting and ending dates of validity as well as conversion factors to euro. For countries not currently using the euro, 2011 exchange rates from Eurostat (2017) were used.

Information on currencies and conversion factors was mostly gathered from ISO 4217 (ISO, 2017) and GHOC databases (Taylor, 2004), supplemented by various Internet resources.

The second table contains deflators used to adjust nominal losses to real losses in 2011 prices. The GDP deflator was generally used, as it allowed us to make the loss adjustments consistent with GDP values. Alternative price indices were used only if the GDP was not available, but they were always "anchored" to the GDP deflator series. These other series included indices of consumer, wholesale, retail, or cost-of-living prices. The source of the data was usually the same as those for the GDP data; they are listed in detail in the data files themselves. 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 data set. The data set also includes deflator series for three former countries – Czechoslovakia, the Soviet Union and Yugoslavia – as many countries were their constituents in the past.

3 Results

3.1 Database contents

The complete list of files of HANZE and their contents is listed in Table 4. Exposure maps in 100 m resolution are provided as GeoTIFF rasters in ETRS89/LAEA projection, consistent with INSPIRE European grid. The baseline maps of land use and population (100 m resolution) are also included. For the benefit of climate research groups in particular, the data sets are provided also in aggregated, lower-resolution versions. Two files in netCDF format are included: 5′ grid in geographical coordinates (WGS84) and finally 0.11° rotated-pole grid as used in EURO-CORDEX climate modelling framework (Jacob et al., 2014).

Input historical statistics and the HANZE-Events database of past damaging floods are provided as Excel files. The structure of the files with input data is detailed in Tables S3 and S4. Apart from the statistical information, the two files (with demographic/environmental and economic data) each includes a table with all sources and transformations made to the data per country, per variable, and per year, as well as a list of references. The contents of the HANZE-Events database with explanations of all data recorded per event is shown in Table 5.

3.2 Data validation

The accuracy of the data involved in HANZE database is influenced by three elements: (1) quality of baseline maps and historical statistics; (2) robustness of the methodologies used for disaggregation of data and modelling change in popula-

tion and land use; and (3) completeness and reliability of the records of past damaging floods.

3.2.1 Baseline maps and historical statistics

The baseline land cover/use map, CORINE Land Cover 2012, was employed for this analysis before final validation was made, but subsequently the map was found to have thematic accuracy of around 90 % (Copernicus Land Monitoring Service, 2017). Still, the use of thresholds of minimum size (25 ha) and width (100 m) of objects necessary for inclusion in the map result in many small objects with large effects on population distribution to be omitted, e.g. small bodies of water or smaller pieces of infrastructure and villages. It should also be noted that mapping was done by country independently, and therefore the classification of land use is not always fully consistent between countries, and the thematic accuracy varied from 82 to 97 % between countries. Validation reports are also available for imperviousness layers and elevation models from Copernicus Land Monitoring Service (2017).

The baseline GEOSTAT population grid's accuracy is described in reports by the provider (Eurostat, 2017). Though for most countries the quality of the 1 km grid is very high, with 98–100 % of a national population georeferenced, there are exceptions. In Bulgaria, for example, only 57 % of the population was georeferenced and the remainder was disaggregated from settlements or local administrative units. In Italy the entire data set was calculated from enumeration areas, albeit their average size was below 1 km². For some smaller countries, the population distribution was calculated by the European Commission's Joint Research Centre using land use data. Basic information on GEOSTAT accuracy per country has been included in the HANZE database.

Historical statistics were compiled from a large variety of sources. Total population figures were mostly available at regional level, while the remaining statistics were usually available only at national level beyond the most recent 2–3 decades. Inevitably, there are inaccuracies from applying national trends at the regional level. Also, economic data series before approx. 1950 for western Europe and 1990 for central Europe are more often than not reconstructions based on ancillary or proxy data. Notwithstanding those limitations, we believe that, for the study area, the HANZE database represents an improvement in resolution and thematic coverage over the HYDE database. A comparison in the number of regional estimates of total and urban population included in both databases is shown in Fig. S12.

3.2.2 Methods

In this study, the population distribution was disaggregated from 1 km to 100 m using two methods validated previously in literature (Batista e Silva et al., 2013). Lack of comparative data at such resolution prevents us from further analysing

Table 4. List of files of HANZE database. XXXX represents the value indicating the year to which data set pertains.

Type	File	Format	Variables/contents
Output	CLC_XXXX	8 bit TIFF	land cover/use type, 44 classes according to CORINE Land Cover
Output	Pop_XXXX	16 bit TIFF	total population per grid cell (in persons)
Output	GDP_XXXX	16 bit TIFF	gross domestic product (GDP) per grid cell per year (× EUR 10 000 in constant 2011 prices)
Output	FA_XXXX	16 bit TIFF	wealth per grid cell (× EUR 100 000 in constant 2011 prices)
Output	Exposure_5min	netCDF	land use (fraction of urban areas, croplands, pastures, forests, water), total population, GDP, and wealth per grid cell aggregated to 5 arcmin resolution
Output	Exposure_cordex_0.11	netCDF	as above, but aggregated to EURO-CORDEX rotated-pole grid, 0.11° resolution
Output	Events_floods	Excel	list of past damaging floods (list of variables in Table 5)
Input	Expo_input_CLC_Pop	Excel	historical land use/cover and population statistics
Input	Expo_input Econ	Excel	historical economic and currency statistics
Input	CLC_base	8 bit TIFF	baseline land cover/use type, 44 classes according to CORINE Land Cover
Input	Pop_base	16 bit TIFF	total baseline (disaggregated) population per 100 m grid cell (in persons)

Table 5. Information included in HANZE-Events database.

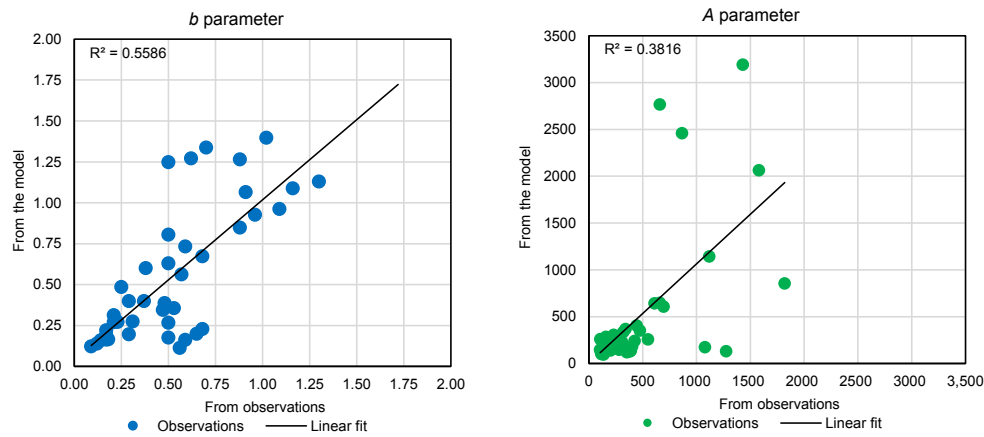| Variable	Description
No.	Event number
Country code	NUTS0 country code
Year	Year of the event (assigned from starting date)
Country name	Country in which the event occurred, using political divisions of the time of the event. In the case of the historical countries of Czechoslovakia, East Germany, USSR, and Yugoslavia, the appropriate successor states were 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, if available. The list of names is usually not comprehensive.
Regions affected	Regions where 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, these were used. If only the number of houses flooded was reported, the number persons affected was estimated by multiplying 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 (millions of 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.

Figure 3. Estimates of A and b parameters (Eq. 10) from modelled and observed population data.

the quality of the disaggregation. Still, the original resolution is very fine and the refining narrows the distribution of population by eliminating areas that are uninhabitable or very unlikely to be inhabited. There is no comparative information for economic variables downscaled from regional level to gridded data.

Lack of comparable data for validation is also evident for historical land use changes. Some local reconstructions of past land cover/use were made from old maps, but there is limited consistency in classification or minimal mapping units to allow for an accurate comparison. CORINE Land Cover is available for 2000 and 2006, but often indicated changes in land use are only reclassifications rather than actual developments. Hence, changes in historical croplands and pasture distribution were not validated directly. The general methodology used here, i.e. reallocating croplands and pastures based on land suitability for agriculture, has been extensively utilized in many studies before (Hurtt et al., 2011; Kaplan et al., 2011; Klein Goldewijk and Verburg, 2013; Pongratz et al., 2008; Ramankutty and Foley, 1999). A more detailed uncertainty and sensitivity analysis of the input data and methods would be possible using structured expert judgment (Colson and Cooke, 2017; Cooke and Goossens, 2008).

Some analysis, however, could be made on the historical distribution on urban population. Estimates with the Clark (1951) model of urban population density are available for 19 cities, which consider population distribution in urban areas as an exponential function:

$$y = Ae^{-bx}, \tag{15}$$

where y is the population density (in persons per km^2), x is the distance from the city centre (in kilometres), and A and b are exponential function coefficients. A total of 42 estimates of this equation spanning a whole century, from 1871 to 1971 were collected, of which a complete list can be found in Table S5 (Clark, 1951, 1967; Hourihan, 1982). In the population map constructed herein, the population density was

calculated for 500 m wide zones around (arbitrarily chosen) city centres, interpolated to match the time points from literature, and then fitted to an exponential function.

A comparison of function parameters is presented in Fig. 3. Overall, a good fit was achieved for the b parameter, but only a relatively poor one for the parameter A. For cities where more than 1 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. (13) parameters would be difficult, since the exponential curve fits are very sensitive to the sample size (distance from the city centre) and the source material: literature studies used census wards of different sizes instead of a disaggregated population grid used here.

Further validation of historical population grid was done by using Eurostat-produced estimates of population at local administrative unit (LAU). This data set (Gløersen and Lüer, 2013) is provided at LAU level 2, except Denmark, Lithuania, Portugal, and Slovenia, where coarser LAU level 1 data are available; data for microstates, except Liechtenstein, are missing. Population is provided at census dates or interpolated/extrapolated to six reference dates (1 January every decade from 1961 to 2011). Data at census dates were extensively used in HANZE database by aggregating them to NUTS3 regions. Here, we connected LAUs in the Eurostat data set with a vector map from Eurostat (2018). For Greece, only LAU level 1 map was available; therefore population estimates were aggregated accordingly. Administrative changes were accounted for to synchronize the population data set and the map, though a small number of LAUs for Ireland and the United Kingdom could not be matched between the data sets (as a result validation was not possible for region UKK14). The final map has 109 177 units, which was then intersected with population grids for 1960, 1970, 1980, 1990, 2000, and 2010. Then, for each LAU two measures commonly used for flood map validation were employed (Alfieri et al., 2014). Test for "correctness" (or "hit rate" I_{cor})
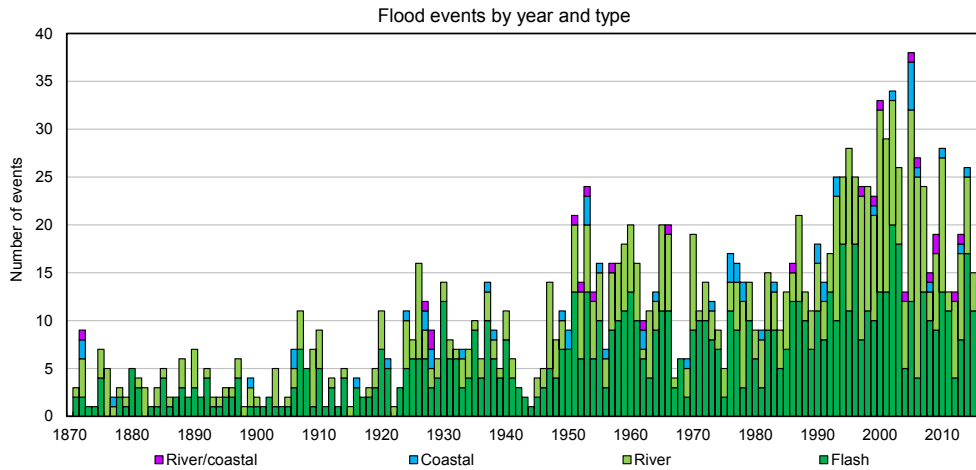
Figure 4. Distribution of flood events in HANZE by year and type.

Table 6. Scores in two measures of accuracy of gridded population estimates (simple average for all LAUs).

Measure	1960	1970	1980	1990	2000	2010
Correctness	88 %	88 %	88 %	90 %	93 %	97 %
Fit	43 %	49 %	57 %	66 %	76 %	93 %

indicates what percentage of the reference unit population is recreated in the HANZE map:

$$I_{cor} = \frac{P_{HM} \cap P_{RM}}{P_{RM}} \cdot 100, \qquad (16)$$

where P_{HM} is the population in the HANZE map and P_{RM} is the population estimate in the reference map. However, this test does not penalize overestimation; therefore another measure for "fit" (or "critical success index" I_{fit}) is applied:

$$I_{fit} = \frac{P_{HM} \cap P_{RM}}{P_{HM} \cup P_{RM}} \cdot 100. \qquad (17)$$

The scores for each NUTS3 region (simple average of LAUs in a given region) can be found in the Supplementary File 2. A simple average of scores for all LAUs is shown in Table 6. The results for 2010 map, which is almost identical to the baseline map, are not 100 % due to relatively low geometrical accuracy of LAU map, use of interpolation in Eurostat data set as opposed to data for the exact year used to produce the population grid, and some missing data at LAU level. Scores for both measures decline over time, and they also vary greatly among countries. It should be noted that population dynamics at the LAU level is significant, as an average LAU changed its population by 91 % between 1960 and 2010 (median change was 36 %), with a decline recorded in 48 % of LAUs

3.2.3 Records of past damaging floods

The quality of records of past floods depends on two main factors: (1) completeness (what share of past floods could be traced) and (2) the reliability of information on the location and quantitative data on losses. Completeness varies substantially between countries, few of which maintain publicly available databases of flood losses. Historical information contained in mandatory preliminary flood risk assessments was sometimes very extensive, but often little or no quantitative information on losses was included. International databases of events have short timespan: EM-DAT nominally starts with year 1900, but very few floods are included before 1970. NatCatService and EEA's compilation of Floods Directive data have coverage from 1980 and Dartmouth Flood Observatory from 1985. Due to the development of Internet, availability of news reports on floods increased substantially starting with mid-1990s, though an increasing number of old newspaper articles are digitized and provide a valuable resource. Under-reporting for central European countries before 1990 is also evident, due to communist-era censorship.

The reliability of past flood loss data remains an open question. Efforts were made to gather multiple sources for past events, especially large ones. In the vast majority of cases, records of floods from international databases can be corroborated by other sources or at least by other international databases. Some records were found to be either dubious or were not primarily flood events, but rather landslides, as found for Portugal (Zêzere et al., 2014). The most extreme case is a record in EM-DAT, according to which a flood along the Danube in Romania in 1926 caused 1000 deaths. However, the Romanian preliminary flood risk assessment indicates that national literature sources do not contain any mention of flood fatalities in that year (Administraţia Naţională Apele Române, 2009). A calamity of such magnitude, which would have been the deadliest European flood in the past 150 years, must have left a trace in several sources. There-

fore, this event was not included in HANZE. Also, there are some cases of floods occurring with other hazards (windstorms, hail, landslides), where it was not possible to disentangle flood losses from those from other causes. Therefore some flood records include or might include those other losses, which are marked in the database under "Notes" category. On the other hand, some flash floods were not included if the majority of losses were not caused by floodwater.

In total, HANZE-Events contains 1564 records of floods (Fig. 4), where 157 events (10 %) have information on the flooded area, 1547 (99 %) on persons killed, 682 (44 %) persons affected, and 560 (36 %) on monetary losses. The known flood consequences amount to almost 123 000 km^2 of land inundated, 18 319 fatalities, 7.5 million people affected, and 227 billion euros in damages in 2011 prices. This can be considered only as a foundation of a comprehensive database. In future work, many more sources of information could be integrated with a larger pan-European effort to overcome the challenges of language barriers and the need to physically access many older sources.

4   Data availability

HANZE-Exposure (both input and output data sets), HANZE-Events and database documentation were uploaded to the 4TU Centre for Research Data (https://doi.org/10.4121/collection:HANZE). Contents of the database were described in Sect. 3.1.

5   Conclusions

The HANZE-Exposure database is intended to provide data allowing one to normalize historical losses related to natural hazards. We hope that it will be useful for researchers studying past occurrences of damaging meteorological, hydrological, or geophysical phenomena. Also, the database could be used to analyse changes in distribution of population and assets within natural hazard zones, e.g. flood hazard maps. To improve reusability, we provide exposure data in different resolutions and formats, so that the data set can be easily applied regardless of how the extent of events ("footprint") is defined: a polygon, a raster layer, a country subdivision, or a climate model grid. HANZE-Exposure can be also considered as a refinement of the HYDE database for Europe for the past 150 years. In principle the spatial data sets and the input historical statistics could also be applied for purely socioeconomic research, e.g. studying regional development or land use changes. However, in that case we would urge potential users to first analyse the methodology and data sources contained in the database and its documentation in order to assess if HANZE-Exposure is suited for the users' research purposes. For example the resolution of regional economic data is of crucial importance when analysing the convergence of the levels of economic development between regions. In

this case, the resolution of regional economic data is of crucial importance Also, the 100 m resolution of the data set should not be interpreted as a benchmark of its accuracy, as it was chosen to (1) preserve the good representation of urban areas and elements of infrastructure, where most of the population lives and most wealth is accumulated and (2) align socioeconomic data with pan-European flood maps which have the same resolution.

HANZE-Events currently encompasses only information on floods, but the same framework could be used for other hazards. Number of casualties or losses in monetary terms can then be corrected (or "normalized") for changes in currency, inflation, population, or economic growth using HANZE-Exposure. Also, reported losses could be contrasted with potential losses (e.g. exposed population or assets within a flood hazard zone with a given probability of occurrence; Paprotny et al., 2017). Information on relative losses could provide insight into how the vulnerability of a population has changed over time.

The Supplement related to this article is available online at https://doi.org/10.5194/essd-10-565-2018-supplement.

References

Administraţia Naţională Apele Române: Studii Privind Determinarea Actiunilor, Masurilor, Optiunilor si Solutiilor Pentru Atingerea Obiectivelor Gestionarii Intergratea a Resurselor

de Apa Aple Bazinelor Hidrografice Districtulde de Bazon Hidrografic Olt, Volumul 6: Managementul Risculuila Inundati, available at: http://www.rowater.ro/daolt/PlanManagement/3.PLANULDEAMENAJAREALB.H.OLT/VOL6.pdf (last access: 17 April 2017), 2009.

Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P., and Feyen, L.: Advances in pan-European flood hazard mapping, Hydrol. Process., 28, 4067–4077, https://doi.org/10.1002/hyp.9947, 2014.

Anas, A., Arnott, R., and Small, K. A.: Urban Spatial Structure, J. Econ. Lit., 36, 1426–1464, 1998.

Barredo, J. I.: Major flood disasters in Europe: 1950–2005, Nat. Hazards, 42, 125–148, 2007.

Barredo, J. I.: Normalised flood losses in Europe: 1970–2006, Nat. Hazards Earth Syst. Sci., 9, 97–104, https://doi.org/10.5194/nhess-9-97-2009, 2009.

Barredo, J. I., Saurí, D., and Llasat, M. C.: Assessing trends in insured losses from floods in Spain 1971–2008, Nat. Hazards Earth Syst. Sci., 12, 1723–1729, https://doi.org/10.5194/nhess-12-1723-2012, 2012.

Batista e Silva, F., Gallego, J., and Lavalle, C.: A high-resolution population grid map for Europe, J. Maps, 9, 16–28, 2013.

Berry, B. L. J., Simmons, J. W., and Tennant, R. J.: Urban Population Densities: Structure and Change, Geogr. Rev., 53, 389–405, 1963.

Black, A. R. and Law, F. M.: Development and utilization of a national web-based chronology of hydrological events, Hydrolog. Sci. J., 49, 237–246, 2004.

Boudou, M., Danière, B., and Lang, M.: Assessing changes in urban flood vulnerability through mapping land use from historical information, Hydrol. Earth Syst. Sci., 20, 161–173, https://doi.org/10.5194/hess-20-161-2016, 2016.

Bouwer, L. M.: Have disaster losses increased due to anthropogenic climate change?, B. Am. Meteorol. Soc., 92, 39–46, 2011.

Bouwer, L. M., Crompton, R. P., Faust, E., Höppe, P., and Pielke, R. A.: Confronting Disaster Losses, Science, 318, 753 pp., 2007.

Brakenridge, G. R.: Global Active Archive of Large Flood Events, Dartmouth Flood Observatory, University of Colorado, available at: http://floodobservatory.colorado.edu/Archives/index.html (last access: 20 May 2017), 2017.

CIESIN/SEDAC: Gridded Population of the World, Version 4 (GPWv4): Population Count, https://doi.org/10.7927/H4X63JVC, last access: 15 January 2018.

Clark, C.: Urban Population Densities, J. R. Stat. Soc. A Stat., 114, 490–496, 1951.

Clark, C.: Population Growth and Land Use, Macmillan, London, 1967.

Colson, A. R. and Cooke, R. M.: Cross validation for the classical model of structured expert judgment, Reliab. Eng. Syst. Safe., 163, 109–120, 2017.

Cooke, R. M. and Goossens, L. L. H. J.: TU Delft expert judgment data base, Reliab. Eng. Syst. Safe., 93, 657–674, 2008.

Copernicus Land Monitoring Service: http://land.copernicus.eu/, last access: 18 July 2017.

Daniell, J. E., Khazai, B., Wenzel, F., and Vervaeck, A.: The CATDAT damaging earthquakes database, Nat. Hazards Earth Syst. Sci., 11, 2235–2251, https://doi.org/10.5194/nhess-11-2235-2011, 2011.

Dirección General de Protección Civil: Catálogo Nacional de Inundaciones Históricas, available at: http://www.proteccioncivil.es/catalogo/naturales/cnih/cnih2014/Presentacion.html (last access: 1 March 2017), 2015.

Domeneghetti, A., Carisi, F., Castellarin, A., and Brath, A.: Evolution of flood risk over large areas: Quantitative assessment for the Po river, J. Hydrol., 527, 809–823, 2015.

Eicher, C. L. and Brewer, C. A.: Dasymetric mapping and areal interpolation: Implementation and evaluation, Cartogr. Geogr. Inf. Sc., 28, 125–138, 2001.

ESRI: Europe Population Density, available at: www.arcgis.com/home/item.html?id=cf3c8303e85748b5bc097cdbb5d39c31, last access: 12 December 2016.

European Environment Agency: Flood phenomena, available at: https://www.eea.europa.eu/data-and-maps/data/european-past-floods/flood-phenomena (last access: 18 December 2016), 2015.

European Union: 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), Official Journal of the European Union, L 13/4, 18.1.2011, 3–54, 2011.

Eurostat: http://ec.europa.eu/eurostat/data/database, last access: 11 July 2017.

Eurostat: GISCO: geographical information and maps, available at: http://ec.europa.eu/eurostat/web/gisco/overview, last access: 15 January 2018.

FAO: Global Agro-Ecological Zones, available at: http://gaez.fao.org/Main.html#, last access: 20 November 2016.

Fischer, G., van Velthuizen, H. T., Shah, M. M., and Nachtergaele, F. O.: Global Agro-ecological Assessment for Agriculture in the 21st Century: Methodology and Results, IIASA Research Report RR-02-02, IIASA, Laxenburg, Austria, 2002.

Fuchs, R., Herold, M., Verburg, P. H., and Clevers, J. G. P. W.: A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in Europe, Biogeosciences, 10, 1543–1559, https://doi.org/10.5194/bg-10-1543-2013, 2013.

Fuchs, R., Verburg, P. H., Clevers, J. G. P. W., and Herold, M.: The potential of old maps and encyclopaedias for reconstructing historic European land cover/use change, Appl. Geogr., 59, 43–55, 2015a.

Fuchs, S., Keiler, M., and Zischg, A.: A spatiotemporal multi-hazard exposure assessment based on property data, Nat. Hazards Earth Syst. Sci., 15, 2127–2142, https://doi.org/10.5194/nhess-15-2127-2015, 2015b.

Gallego, F. J.: A population density grid of the European Union, Popul. Environ., 31, 460–473, 2010.

Gallego, F. J., Batista, F., Rocha, C., and Mubareka, S.: Disaggregating population density of the European Union with CORINE land cover, Int. J. Geogr. Inf. Sci., 25, 2051–2069, 2011.

Gisgraphy: http://download.gisgraphy.com/openstreetmap/pbf/, last access: 20 August 2017.

Gløersen, E. and Lüer, C.: Population data collection for European local administrative units from 1960 onwards – final report, Spatial Foresight GmbH, Luxembourg, 17 pp., 2013.

Guha-Sapir, D., Below, R., and Hoyois, P.: EM-DAT: The CRED/OFDA International Disaster Database, Université

Catholique de Louvain, Brussels, Belgium, available at: http://www.emdat.be/, last access: 23 April 2017.

Guzzetti, F. and Tonelli, G.: Information system on hydrological and geomorphological catastrophes in Italy (SICI): a tool for managing landslide and flood hazards, Nat. Hazards Earth Syst. Sci., 4, 213–232, https://doi.org/10.5194/nhess-4-213-2004, 2004.

Haigh, I. D., Wadey, M. P., Gallop, S. L., Loehr, H., Nicholls, R. J., Horsburgh, K., Brown, J. M., and Bradshaw, E.: A user-friendly database of coastal flooding in the United Kingdom from 1915–2014, Scientific Data, 2, 150021, https://doi.org/10.1038/sdata.2015.21, 2015.

Hourihan, K.: Urban Population Density Patterns and Change in Ireland, 1901–1979, Econ. Soc. Rev., 13, 125–147, 1982.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., K., Klein Goldewijk, Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P., and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Climatic Change, 109, 117–161, 2011.

ISO: Codes for the representation of currencies, ISO 4217:2015, available at: https://www.currency-iso.org/en/home.html, last access: 11 May 2017.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Reg. Environ. Change, 14, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.

Joint Research Centre: Global Land Cover 2000, available at: http://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php (last access: 16 November 2016), 2015.

Jongman, B., Ward, P. J., and Aerts, J. C. J. H.: Global exposure to river and coastal flooding: Long term trends and changes, Global Environ. Chang., 22, 823–835, 2012.

Jongman, B., Koks, E. E., Husby, T. G., and Ward, P. J.: Increasing flood exposure in the Netherlands: implications for risk financing, Nat. Hazards Earth Syst. Sci., 14, 1245–1255, https://doi.org/10.5194/nhess-14-1245-2014, 2014.

Kaplan, J. O., Krumhardt, K. M., and Zimmermann, N. E.: The effects of land use and climate change on the carbon cycle of Europe over the past 500 years, Glob. Change Biol., 18, 902–914, 2011.

Klein Goldewijk, K. and Verburg, P. H.: Uncertainties in global-scale reconstructions of historical land use: an illustration using the HYDE data set, Landscape Ecol., 28, 861–877, 2013.

Klein Goldewijk, K., Beusen, A., van Drecht, G., and de Vos, M.: The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years, Global Ecol. Biogeogr., 20, 73–86, 2011.

Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land use estimates for the Holocene – HYDE 3.2, Earth Syst. Sci. Data, 9, 927–953, https://doi.org/10.5194/essd-9-927-2017, 2017.

Kummu, M., de Moel, H., Salvucci, G., Viviroli, D., Ward, P. J., and Varis, O.: Over the hills and further away from coast: global geospatial patterns of human and environment over the 20th–21st centuries, Environ. Res. Lett., 11, 034010, 2016.

Lang, M., Coeur, D., Audouard, A., Villanova-Oliver, M., and Pène J.-P.: BDHI: a French national database on historical floods, E3S Web Conf., 7, 04010, https://doi.org/10.1051/e3sconf/20160704010, 2016.

Lehner, B., Reidy Liermann, C., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J., Rödel, R., Sindorf, N., and Wisser, D.: High resolution mapping of the world's reservoirs and dams for sustainable river flow management, Front. Ecol. Environ., 9, 494–502, https://doi.org/10.1890/100125, 2011.

Milego, R. and Ramos, M. J.: Disaggregation of socioeconomic data into a regular grid and combination with other types of data, Technical Report, ESPON EGTC, Luxembourg, available at: https://www.espon.eu/sites/default/files/attachments/2.2_TR_grids.pdf (last access: 6 March 2018), 2011.

Munich Re: Innovative new ways of analysing historical loss events, https://www.munichre.com/topics-online/en/2016/topicsgeo2015/historical-loss-events-revisited (last access: 16 July 2017), 2016.

Munich Re: https://www.munichre.com/en/reinsurance/business/non-life/natcatservice/index.html, last access: 8 August 2017.

Murakami, D. and Yamagata, Y.: Estimation of gridded population and GDP scenarios with spatially explicit statistical downscaling, arXiv, 1610.09041, 2017.

Nordhaus, W. and Xi, C.: Geographically based Economic data (G-Econ), available at: http://gecon.yale.edu/ (last access: 27 November 2016), 2011.

OurAirports: http://ourairports.com/data/, last access: 15 June 2017.

Papageorgiou, Y. Y.: Population density in a central-place system, J. Regional Sci., 54, 450–461, 2014.

Paprotny, D., Morales-Nápoles, O., and Jonkman, S. N.: Efficient pan-European river flood hazard modelling through a combination of statistical and physical models, Nat. Hazards Earth Syst. Sci., 17, 1267–1283, https://doi.org/10.5194/nhess-17-1267-2017, 2017.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A. J., Freire, S., Halkia, S., Julea, A., Kemper, T., Soille, P., and Syrris, V.: Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014, JRC Technical Report EUR 27741 EN, https://doi.org/10.2788/253582, 2016.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, Global Biogeochem. Cy. 22, GB3018, https://doi.org/10.1029/2007GB003153, 2008.

Ramankutty, N. and Foley, J. A.: Estimating historical changes in global land cover: croplands from 1700 to 1992, Global Biogeochem. Cy., 13, 997–1027, 1999.

Röthlisberger, V., Zischg, A., and Keiler, M.: Spatiotemporal aspects of flood exposure in Switzerland, E3S Web Conf., 7, 08008, https://doi.org/10.1051/e3sconf/20160708008, 2016.

Schiermeier, Q.: Insurers' disaster files suggest climate is culprit, Nature, 441, 674–675, 2006.

Schumacher, I. and Strobl, E.: Economic development and losses due to natural disasters: The role of hazard exposure, Ecol. Econ., 72, 97–105, 2011.

Sofia, G., Roder, G., Dalla Fontana, G., and Tarolli, P.: Flood dynamics in urbanised landscapes: 100 years of climate and humans' interaction, Scientific Reports, 7, 40527, https://doi.org/10.1038/srep40527, 2017.

Stevens, A. J., Clarke, D., Nicholls, R. J., and Wadey, M. P.: Estimating the long-term historic evolution of exposure to flooding of coastal populations, Nat. Hazards Earth Syst. Sci., 15, 1215–1229, https://doi.org/10.5194/nhess-15-1215-2015, 2015.

Stevens, A. J., Clarke, D., and Nicholls, R. J.: Trends in reported flooding in the UK: 1884–2013, Hydrolog. Sci. J., 61, 50–63, 2016.

Taylor, B.: Global History of Currencies, available at: https://www.globalfinancialdata.com/GFD/Article/global-history-of-currencies (last access: 6 March 2018), 2004.

United Nations: World Urbanization Prospects, available at: https://esa.un.org/unpd/wup/ (last access: 10 December 2016), 2014.

Zêzere, J. L., Pereira, S., Tavares, A. O., Bateira, C., Trigo, R. M., Quaresma, I., Santos, P. P., Santos, M., and Verde, J.: DISASTER: a GIS database on hydro-geomorphologic disasters in Portugal, Nat. Hazards, 72, 503–532, 2014.

www.earth-syst-sci-data.net/10/565/2018/

Earth Syst. Sci. Data, 10, 565–581, 2018