

Further validation of historical population grid was done by using Eurostat-produced estimates of population at local administrative unit (LAU). This dataset (Gløersen and Lür 2013) is provided at LAU level 2, except Denmark, Lithuania, Portugal and Slovenia, where coarser LAU level 1 data is available; data for microstates, except Liechtenstein, is missing. Population is provided at census dates or interpolated/extrapolated to six reference dates (1<sup>st</sup> January every decade from 1961 to 2011). Data at census dates was extensively used in the HANZE database, by aggregating them to NUTS3 regions. Here, we connected LAUs in the Eurostat dataset 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 dataset and the map, though a small number of LAUs for Ireland and United Kingdom could not be matched between the datasets (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”  $I_{cor}$ ) indicates what percentage of the reference unit population is recreated in the HANZE map:

$$I_{cor} = \frac{P_{HM} \cap P_{RM}}{P_{RM}} \times 100 \quad (16)$$

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:

$$I_{fit} = \frac{P_{HM} \cap P_{RM}}{P_{HM} \cup P_{RM}} \times 100 \quad (17)$$

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 the 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 dataset 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 as we move further in the past; they also vary a lot between countries. It should be noted that the dynamics of population at 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

**Table 1. 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%