

Deliverable D1.5

Report on methodology to project LU-LC in a given area based on the analysis of existing remote sensing data and AI

Revision	Organization	Date
Written by	DLR	09-09-2024
Checked and approved by	EURAC & IIASA	28-11-2024
Validated and released by	UT	30-11-2024



Information table

Deliverable Number	D1.5
Deliverable Title	Report on Methodology to project LU-LC in a given area based on the analysis of existing remote sensing data and AI
Version	V3
Status	Final
Responsible Partner	Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)
Contributors	Christian Geiß, Lea Schollerer, Vanessa Rittlinger (DLR)
Contractual Date of Delivery	30-11-2024
Actual Date of Delivery	30-11-2024
Dissemination Level	PUB (NOT Public until publication of scientific paper)

Abstract

This Deliverable D1.5 focusses on the PARATUS methodology how Earth observation (EO) datasets together with new Artificial Intelligence (AI) techniques offer innovative possibilities to map future exposure of people and elements to natural hazards in the future. In order to develop mitigation strategies for possible future damage events, this detailed information on the future spatial distribution of the population and further exposed elements are required. Current EO datasets, in particular long time series data with high temporal and thematic resolution, and new techniques from the field of AI, like Long Short-Term Memory Cells, offer innovative possibilities to extrapolate exposure information spatiotemporally. Specifically, EO time series data is compiled that describe changes in global population and land use since around 2000, while providing high spatial, temporal, and thematic resolution. The different datasets are preprocessed to the same temporal (years 2000 – 2020) and spatial extent (100 m pixels). In combination with static features (Fig. 6 to 9) the time series then serves as the basis for a novel AI-model, that identifies characteristic change trajectories in the target variables over time and can extrapolate the target variables correspondingly in spatiotemporal terms into the future. This can be done for instance for the highly dynamic megacity of Istanbul, which is exposed to earthquakes and landslides. The resulting exposure datasets can then be used for an early and sustainable urban planning, risk assessment, and risk reduction efforts in the future, as well as for evaluating the systemic risk and vulnerability of human settlements.



Co-funded by
the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation program under grant agreement No 101073954

Document history

Version	Date	Author	Description
V1	02-09-2024	Lea Schollerer (DLR)	Document creation and adding information about exposure modelling
V1	09-09-2024	Lea Schollerer (DLR)	Including information and figures about the data and methodology
V1	16-09-2024	Vanessa Rittlinger (DLR)	Including general information about the PARATUS Project, WP1, Task 1.3 and D1.5
V1	14-10-2024	Lea Schollerer, Vanessa Rittlinger (DLR)	Adding the content to the experimental setup, results and outlook
V1	11-11-2024	Lea Schollerer, Vanessa Rittlinger (DLR)	Adding of further details, figures, revision of content and general information
V2	19-11-2024	Christian Geiß (DLR)	Internal Review on partner's side
V2	21-11-2024	Massimiliano Pittore (EURAC), Tatiana Ermolieva, Nadejda Komendantova (IIASA)	Internal Review
V3	28-11-2024	Christian Geiß, Vanessa Rittlinger, Lea Schollerer (DLR)	Addressing the reviewer comments and adapting the deliverable
V3	29-11-2024	Cees van Westen (UT)	Submitted the deliverable

Disclosure Statement:

The information contained in this document is the property of the PARATUS Consortium and should not be reproduced, disclosed, modified, or communicated to any third parties without the prior written consent of the abovementioned entities.

The deliverable 1.5 submitted to the European Commission on 30/11/2024 and is waiting for approval by the Research Executive Agency. Therefore, this current version may not represent the final version of the deliverable.

About PARATUS:

The PARATUS project aims to increase the preparedness of first and second responders in the face of multi-hazard events and to reduce the impacts on various sectors resulting from complex disasters. The outcome is to develop an open-source cloud-based Online Service Platform that offers support in reducing dynamic risk scenarios and systemic vulnerability caused by multi-hazard disasters. To achieve these objectives, the project will perform in-depth assessments of complex interactions between hazards and their resulting impacts on various sectors, analyse the current risk situation and study how alternative future scenarios could change multi-hazard impact chains. Based on these analysis, scenarios of multi-hazard impacts will be co-designed and developed with stakeholders in four case study areas (including the Caribbean, Romania, Istanbul, and Alpine regions).



Co-funded by
the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation program under grant agreement No 101073954

List of Acronyms

Acronym	Definition
AI	Artificial Intelligence
ANN	Artificial neural network
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long Short-Term Memory
CORINE	Coordination of Information on the Environment program
CLC	Corine Land Cover
CLMS	Copernicus Land Monitoring Service
DEM	Digital Elevation Model
EEA	European Environment Agency
DLR	German Aerospace Center
JRC	Joint Research Center
LSTM	Long Short-Term Memory
LULC	Land Use and Land Cover
ML	Machine Learning
MTL	Multitask learning
OSM	OpenStreetMap
PARATUS	Increasing Preparedness and Resilience of European Communities by Co-Developing Services Using Dynamic Systemic Risk Assessment
RF	Random forest
RMSE	Root-mean-squared error
RNN	Recurrent neural network
SLEUTH	Slope, Land cover, Excluded regions, Urban land cover, Transportation, and Hillshade
STL	Single task learning
WP	Work Package
WSF	World Settlement Footprint

Executive summary

Work Package 1 *“Learning from the past: understanding the dynamic and interactive conditions of risk”* focuses on utilizing information from historical disaster events (so-called Learning case studies) and combining this information with disaster history in application case studies. This includes analysing how the hazardous events directly and indirectly impacted and would impact different sectors, such as health, cultural heritage, environment and biodiversity, public finance, and key economic sectors. Learning from past events allows us to advance risk science and achieve improved multi-hazard impact forecasting.

This Deliverable D1.5. aims to develop a methodology to project Land Cover and Land Use (LULC) in a given area based on the analysis of existing remote sensing data and artificial intelligence (AI), within the EU Horizon Europe PARATUS project¹. Remote sensing time series data, that represent the trajectory of a change in the past till the present combined with advanced artificial intelligence techniques, is exploited to provide risk components for systemic risk assessment. The data and methods are used to develop indicators for precursors for different types of threats, and monitoring of the risk indicators as input data for systemic risk assessment.

The created multi-temporal exposure datasets, like land use, and population, are future, scenario-based exposure modelling and are used to support risk assessment. Within the project the aim is to provide a complete picture of risk by modelling and anticipating the exposure and vulnerability of the built environment in addition to population forecasts. The resulting exposure assessment datasets can then be used for early and sustainable urban planning, risk assessment, and risk reduction efforts in the future, as well as for evaluating the systemic risk and vulnerability of human settlements.

This deliverable reports on the use of monitoring data and remote sensing services for exposure modelling. In PARATUS this scenario-based exposure modelling is used as input into the understanding of impact chains and systemic risk and for impact assessment (see WP2). In the context of this deliverable the approach has been applied in the application case of Istanbul. Within PARATUS the application case studies serve the purpose of testing the methods on real test cases, working with local practitioners, and simultaneously enhancing and transferring the impact of the research. Istanbul has been chosen as test case to develop and test the methodology since it is a highly dynamic megacity, which is greatly exposed to earthquakes, heatwaves, floods and landslides. In the following steps of the project the methodology should also be applied in other test cases like Bucharest (Romania). Additionally, the methodology itself should be further improved and other suitable datasets can be included as well. These advancements will help strengthen systemic risk assessment methodologies and enhance the practical utility of the findings in disaster risk reduction efforts.

¹ <https://www.paratus-project.eu/>

Table of Contents

DOCUMENT HISTORY.....	3
LIST OF ACRONYMS	4
EXECUTIVE SUMMARY	5
TABLE OF CONTENTS	6
1. INTRODUCTION TO WP1 AND D1.5	7
2. INTRODUCTION TO EXPOSURE MODELLING	11
3. MATERIAL AND METHODS.....	13
3.1. STUDY AREA	13
3.2. TIME-VARIANT DATASETS.....	14
3.2.1. <i>Population</i>	14
3.2.2. <i>Land use land cover</i>	15
3.2.3. <i>World Settlement Footprint (WSF) Evolution</i>	16
3.3. DRIVING FACTORS	16
3.4. MULTITASK LONG SHORT-TERM MEMORY MODEL	18
4. EXPERIMENTAL SETUP	19
5. EXPERIMENTAL RESULTS AND OUTLOOK	20
6. REFERENCES	24

1. Introduction to WP1 and D1.5

Work Package 1 (WP1) *“Learning from the past: understanding the dynamic and interactive conditions of risk”* focuses on utilizing information from historical disaster events (so-called Learning case studies) and combining this information with disaster history in application case studies. This includes analysing how the hazardous events directly and indirectly impacted and would impact different sectors, such as health, cultural heritage, environment and biodiversity, public finance, and key economic sectors. Learning from past events allows us to advance risk science and achieve improved multi-hazard impact forecasting. Moreover, WP1 activities include the exploitation of satellite remote sensing data for providing risk components for systemic risk assessment.

Hence, PARATUS methodologies (and therefore WP1-related work) incorporate two distinct types of case study approaches: application case studies and learning case studies. Learning case studies (see Table 1.1) encompass a collection of previous disaster events aimed at extracting valuable insights. By analysing historical data, these case studies provide a means to comprehend the effects of past disasters on various sectors. On the other hand, application case studies (see Table 1.2) serve the purpose of testing the methods on real test cases, working with local practitioners, and simultaneously enhancing and transferring the impact of the research.

Table 1.1. Learning Case Studies of the PARATUS project

Learning Case Studies (European)	
2021 European floods	Extreme rainfall event leading to severe flooding and heavy damage in different sectors in Germany, Belgium and Netherlands: relevant cross-border disaster example.
2021 Southern European heatwave and wildfires	Severe impact on Greece, Italy and Turkey - interesting because of the multi-sectoral impacts (also touches on cultural heritage sites, vulnerable groups being more affected by heat).
2020 Gloria storm	Catastrophic winter storm that affected NE Spain and S France by winds and heavy rainfall. Severe impacts to coast, hillslopes and floodplains due to different cascading and compounding events.
2020-2021 COVID-19 pandemic	We will look specifically how during the COVID-19 crisis other compounding events caused larger impact due to COVID-19 restrictions and large pressure on health system in various locations, i.e. Romania (floods), Sint Maarten etc.
Landslides observatory, Lower Austria	In 2014 the long-term landslide observatory “NoeSLIDE – Monitoring of different landslide types in Lower Austria”, a multi-parameter measuring network for landslide monitoring was initiated.
2018 VAIA storm	Impact to forests, buildings, roads, due to cascaded hazards (landslides, fluvial floods)
2017 Irma and Maria	Severe Hurricane impact on St. Maarten and other Caribbean islands.
2017 Space weather event	September 2017's geoeffective space weather and impacts to Caribbean radio communications during hurricane response.
2010 Eyjafjallajökull Volcanic Eruption	Volcanic eruption and ash cloud causing severe disruption of air space. The event sets an example to the global impact of a local disaster.

2005 Flood Disaster in Tyrol/Austria	The 2005 European floods hit mainly Romania, Switzerland, Austria and Germany during August 2005. Based on this event, a regional study was carried out in Tyrol, which also considered the road and rail infrastructure.
Earthquake August 17 th 1999 Kocaeli and February 6 th 2023 Kahramanmaraş Earthquakes in Turkey	After the Izmit Earthquake, the national disaster risk management system has been improved by the means of new legal tools and organizations. Since 1999 several earthquakes occurred but they were not as devastating as Kahramanmaraş Earthquakes on 6 th of February 2023. These two earthquakes are important to see the differences between planned actions and the reality due to cascading events.
Vrancea earthquakes, Romania	Historical earthquakes. Cross-border impact by several earthquakes in the Vrancea region, Romania in the 20th and 21st century affecting Bucharest, the most earthquake-prone capital in Europe.
Learning Case Studies (Global)	
August 2021 Haiti	Tropical storm hits after the earthquake in a region with political unrest.
Montserrat & St. Vincent and the Grenadines (SVG), Barbados in April 2021	Strong explosions generated ash plumes, which are not only impacting the entire island of Saint Vincent but also transporting massive amounts of ash to the neighbouring island country of Barbados in the Lesser Antilles of the West Indies. A team of UNU-GLOMOS led a Remote Environmental Assessment & Analysis Cell to support the UN mission teams.
India: Chamoli District on 7 February 2021	The flood swept away the unfinished Tapovan Vishnugad Hydropower Project and inflicted substantial damage on the Rishi Ganga Hydropower Project. Even as rescue and relief efforts were underway, a glacial lake outburst flood (GLOF) was suggested as the reason for the flood, possibly triggered by glacier collapse (Info source: ICIMOD: UNU-EHS has MoU).
India: Kerala Floods and Landslides in 2018	This historic flood led to the conduct of India's first PDNA exercise and one of India's most expensive reconstruction projects.
2011 Tohoku Earthquake and Tsunami and 1995 Great Hanshin Earthquake	The 2011 event provides data on complex reactions and dynamic situations during a disaster. On the other hand, the 1995 Great Hanshin Earthquake was the first earthquake in a metropolitan city. Legacies are still valid. We will investigate the improvements in DRR and DRM from 1995 until 2011 in Japan. UT is in contact with the Disaster Prevention Centre at Kyoto University. UNU-EHS has an agreement with Tohoku University.
2005 Hurricane Katrina	Heavy rainfall and storm surge caused extensive flooding of a major part of New Orleans. Major social and economic disruption, massive outmigration of population. Recovery dynamics.

Table 1.2. Application Case Studies of the PARATUS project

Application Case Studies
Turkey: Istanbul
Romania: Bucharest
Caribbean: Sint Maarten & Saint Vincent & Regional scale
Alps: Brenner corridor

Both Learning and Application Case Studies have been selected due to their unique characteristics concerning stakeholders, hazard interactions, impacts on different combinations of sectors, their diverse vulnerabilities, scale, and planning decisions. The coordination of Application Case Studies is part of Work Package 6; this includes the application and integration of activities belonging to different WPs, encompassing the involvement of the stakeholders. Therefore, more details on the different case studies can be found in the deliverables of WP6. Specifically, deliverable D6.1 “Strategy and Case Study Protocols” describes the strategies and case study protocols to be applied in four main Application Case Study areas. Deliverable D6.2 “Report 1 of Workshops in Application Case Studies” and Deliverable D6.3 “Report 2 of Workshops in Application Case Studies” provide an overview of the workshops that have been organized in the period from January 2023 to March 2024 in the four Application Case Studies and present the main findings.

In WP1, the elected events are studied by developing Impact Chains with the help of Forensic Analysis, remote sensing, and disaster databases. The final objective of this work is to achieve a qualitative and quantitative conceptualization of systemic risk related to complex disaster events.

In total WP1 includes 6 deliverables (see Table 1.3).

Table 1.3: The list of deliverables in WP1

#	Name	Due date (month)	Description
D1.1	Report on participatory workshops in the four Application Case Study sites, including Impact Chains diagram for each analysed event.	10	Lead Beneficiary: EURAC
D1.2	Report on virtual participatory workshops in all learning and Application Case Study sites, including Impact Chains diagram	20	Lead Beneficiary: EURAC
D1.3	Impact, damage and loss data related to the considered events in the four core cases. Wiki-type structured Database.	20	Lead Beneficiary: EURAC
D1.4	Guidelines on the actionable assessment of hazards interactions in compound / cascading events	24	Lead Beneficiary: UNIVIE
D1.5	Methodology to project LU-LC in a given area based on the analysis of existing remote sensing data and AI	24	Lead Beneficiary: DLR
D1.6	Report on the exemplification of the use of remote sensing services analysis solutions	36	Lead Beneficiary: DLR

The deliverable D1.5 with the focus on exploitation of satellite remote sensing data and artificial intelligence for providing risk components for systemic risk assessment aims to develop indicators for precursors for different types of threats and monitoring of risk indicators. Therefore, monitoring data and remote sensing services are integrated to map impacts by historical disaster events. This approach is an input into the understanding of impact chains and systemic risk by using satellite-based extended time series and AI approaches integrated with other data sources to provide multi-temporal exposure datasets (e.g., roads,



Deliverable 1.5. Report on Methodology to project LU-LC in a given area based on the analysis of existing remote sensing data and AI

buildings and affiliated hazard-related structural types, critical infrastructure, land use, and population) for future, scenario-based exposure modelling to support risk assessment. Therefore, it is input for other work packages, like WP2 that is looking ahead and focussing on today's future by doing systemic risk assessment for current situation but also developing scenarios for future change.

After this introduction which summarises the main links to the PARATUS project, this deliverable provides a short introduction on exposure modelling (Chapter 2) followed by a description of the used materials and the developed method (Chapter 3). Chapter 4 and 5 shows the experimental setup, the achieved results and gives also an outlook on future work.



Co-funded by
the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation program under grant agreement No 101073954

2. Introduction to exposure modelling

There has been an increase in the number of recorded natural hazard events in the last decades (Ritchie et al., 2024a). Such events can cause huge losses, especially in human settlements with high population densities. It can be expected that this situation intensifies in the future as the world's population is still growing within the coming 50 to 60 years to an estimated number of 10.3 billion people (UN, 2024). Further, the ongoing climate change triggers both single and multi-hazard disasters with a higher risk of more extreme events (Ritchie et al., 2024b). As a result, more people will be exposed to natural hazards in the future than ever before. In order to develop mitigation strategies for possible future damage events, the understanding and modelling of future risks from natural hazards is getting more crucial and because of that detailed information on the future spatial distribution of population and further exposed elements are required (Cremen et al., 2022). This urge is also recognised by the Sendai Framework for Disaster Risk Reduction (2015-2030) and noted in the United Nations 2030 Agenda for Sustainable Development. This emphasises the need to prepare for disasters in our world, and especially urban areas, since more than 50% of growth until 2025 is expected to happen in cities (Stevens et al., 2015). Resulting strategies are towards the minimization of uncontrolled development in hazard-prone areas. While many natural-hazard risk-assessment frameworks focus more on the current conditions and the historical context more researchers also provide risk-quantification approaches for a sustainable future (Cremen et al., 2022). To obtain this information the aim is to provide risk components about future population and land use land cover change by using a novel spatio-temporal deep learning approach applied to remote sensing-based time series data.

The need for data on human population distributions is driven by a number of factors, including the necessity of measuring impacts of population growth, monitoring human-environment interactions and the requirement for urban planning (Stevens et al., 2015). In response to this, a number of methods have already been developed to predict future population distribution spatially based on explicitly defined rules (Geiß et al., 2024). Two popular techniques are cellular automata (CA) and agent-based models (ABM). Both techniques examine population growth based on local rules and the interaction of neighbouring cells or agents (Clarke, 2014). Furthermore, there are Markov chain models as stochastic ones, that describe a potential sequence of events or states, with the probability dependent on the previous state (Gagniuc, 2017). These methods are often based on simple rules that have to be defined with expert knowledge, which can cause biases in the simulation (Maithani, 2009). While this simplicity is beneficial in some respects, it can also result in the loss of important details (Zhu et al., 2024). This is a reason, why Machine Learning (ML) is increasingly being used to forecast population change. The approach involves deriving a decision rule from prior knowledge, such as a time series (Zhu, 2022) and so to say “learns from the past to predict a future situation”. This approach and used datasets rely on a data driven method, that is trained by utilizing spatiotemporal information. For instance, Chen et al. (2020) predicted residential land use and population distribution with decision-tree-based models. Kubota et al. (2022) developed a Convolutional Neural Network (CNN) to predict short-term population change by using mobile device information. Zheng and Zhang (2020) used social media geolocation data to predict weekly population distribution in China. They employed a Convolutional LSTM (ConvLSTM) model to achieve this. Also, by utilizing LSTM-based models,



Geiß et al. (2024) published a method to forecast long-term population development using observation (EO) derived features. This work is based on the same approach.

The adaptation of land cover to human needs is leading to its increased use in different ways, which in turn can increase exposure to natural hazards (Krüger, 2016). In the context of natural-hazard risk assessment, the focus of land use change is on urban growth (Cremen et al., 2022). In general, human-induced LULC changes are challenging to predict. To gain insight into these complex processes models based on time series and historical data are often employed to forecast future trends (Krüger, 2016; Marconcini et al., 2021). Therefore, analogous methodologies to those employed in forecasting population have been implemented. For instance, Cellular Automata, Agent-based and Markov chain models. One well-known approach is the SLEUTH (Slope, Land cover, Excluded regions, Urban land cover, Transportation, and Hillshade) model (Clarke, 2008). As aforementioned, instead of rule-based methods there is a growing application of Deep Learning (DL). For instance, Qian et al. (2020) combined different techniques and used a 3D CNN to extract spatiotemporal neighbourhood features from land use data in order to predict land use change. To simulate urban expansion Pan et al. (2022) applied a LSTM together with a U-Net. Recently, transformer-based models have also been employed in a variety of fields. (Zhu et al., 2024) utilised one such model to simulate urban expansion at an intra-urban granular level within an explainable ensemble deep learning framework.

Inspired by human learning, it is often beneficial to combine related tasks and learn them jointly, as this allows for the leveraging of shared information. Multitask learning (MTL) methods are used to enhance generalisation performance for more effective representation learning (Zhang and Yang, 2022). As a result, MTL is becoming increasingly popular in a number of different fields, like meteorology to forecast weather (Yan et al., 2024), in urban and smart city planning to forecast traffic flow (Karimzadeh et al., 2021), in natural language modelling for sentiment classification (Jin et al., 2020) or in medicine for surgeries (Islam et al., 2021). It has also been employed in the risk assessment context, for example to classifying multicriteria street-level images with five vulnerability-related target variables (Aravena Pelizari et al., 2023). Given the strong correlation between population distribution and land cover types, and the subsequent impact on spatial population distribution and land cover types (Stevens et al., 2015), within the Task 1.3 a DL multitask learning (MTL) approach was developed to simultaneously predict population and land use and land cover (LULC) based on past changes. Based on the combination of information this approach seems to be suitable within the PARATUS project focusing on cascading natural hazard events and the analysis of the events from multiple directions.

3. Material and Methods

3.1. Study area

The developed method is tested and applied in one of the highly relevant test case areas of the PARATUS project: the megacity of Istanbul in Turkey. Due to its high exposure to natural hazards, like heatwaves, droughts, floods and especially earthquakes and landslides, it represents a critical case study (Portoghese, 2024). Istanbul experienced a rapid population growth in the last two centuries from 8 to 16 million people, increasing the pressure on the city's infrastructure and resources. The largest city of Turkey is located near the North Anatolian Fault (NAF), one of the most seismically active and dangerous fault lines in the world. The city also has a history of devastating earthquakes, with the 1999 Kocaeli (Izmit) earthquake as reminder of the region's seismic vulnerability. Further the probability of a significant earthquake occurring in the next 30 years after the 1999 event is estimated at 35-70% (Parsons, 2004; Kundak et al., 2024).

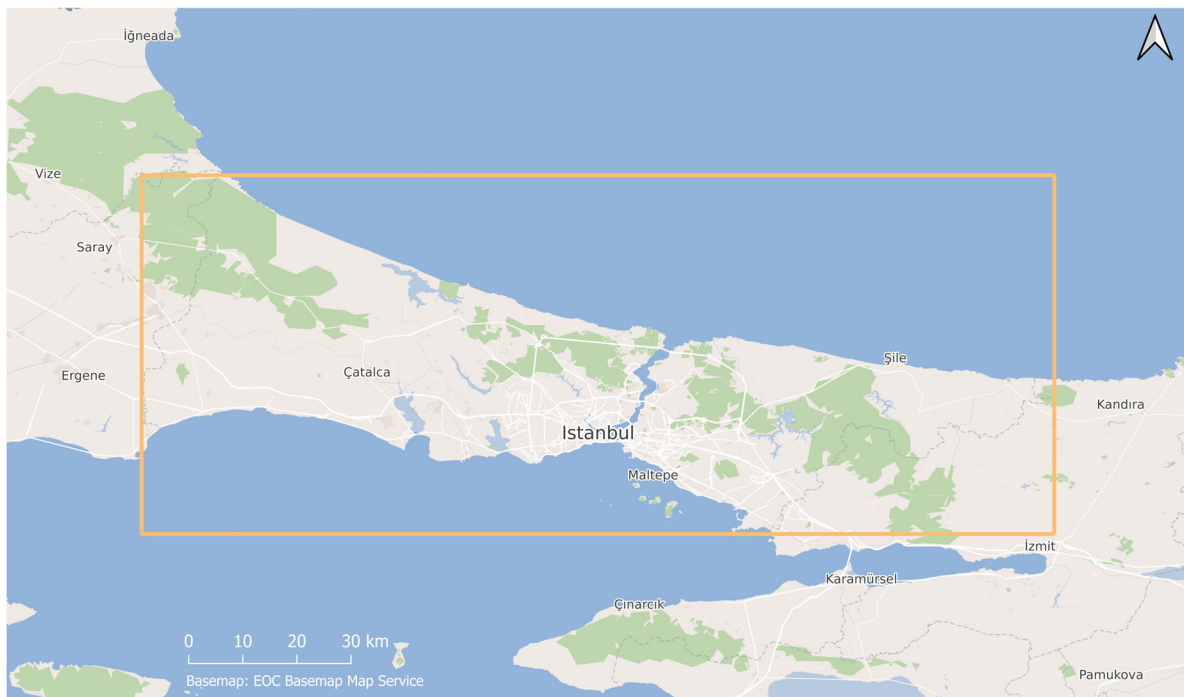


Figure 1: Boundary of the case study area of Istanbul

The spatial extent of the test area (Figure 1) was derived from the administrative boundary provided by OpenStreetMap (OSM). The selected extent covers the full metropolitan area, including both the European and Asian sides of the city and thereby capturing the diverse range of urban, peri-urban, and rural landscapes that contribute to the city's overall vulnerability profile. By using the OSM boundaries a free and open available dataset has been chosen, as well as to keep consistency in the spatial delineation with regard to other study areas.

3.2. Time-variant datasets

Following an exhaustive investigation of data pertaining to exposure, which was also linked to the input provided for deliverable D4.1, two multitemporal gridded datasets were harvested, harmonised and aligned to a set of geospatial covariates, i.e. driving factors, in order to map the trajectory of past development. The aim was to have a collection of freely and openly accessible remote-sensing based datasets with a high spatial, temporal and timely resolution, and a coverage for Europe or even worldwide. All datasets are harmonized and aligned to the raster grid of the population data.

The different datasets are described in more detail below.

3.2.1. Population

To represent the spatiotemporal population change, the gridded multitemporal population distribution of the WorldPop initiative is used² (Figure 2). The raster data set includes the population counts, with the absolute number of people in that area in a 100-meter grid, beginning in 2000 and continuing until 2020 on a yearly basis. Furthermore, the population counts are adjusted to align with the United Nations population estimates³. For the generation of the WorldPop dataset a Random Forest (RF)-based semi-automated dasymetric modelling approach is applied that includes detailed census and ancillary data. In order to create the dasymetric weights, a variety of remote sensing and geospatial data is incorporated, including country-specific census counts linked to corresponding administrative units, landcover, elevation and slope, road networks and waterways, night-time lights, large water bodies, settlement or populated place locations, climate data, protected area delineations, and various “facility” locations. This is done prior to the gridded

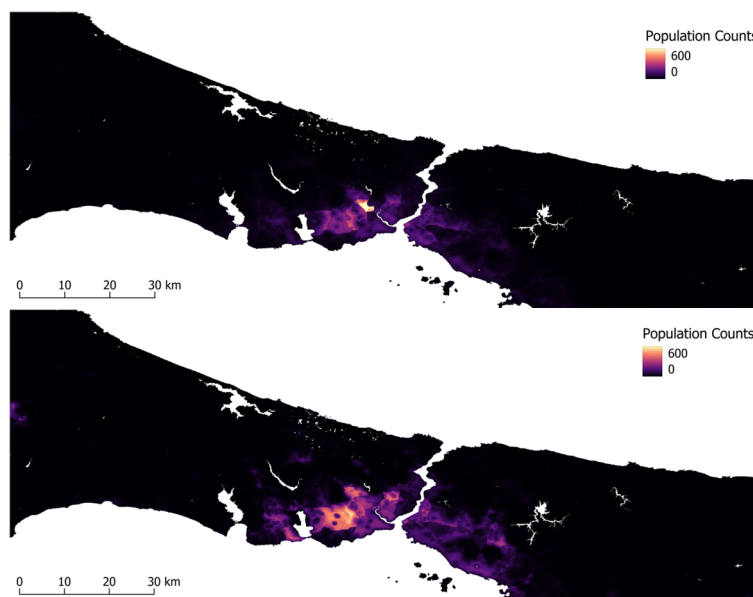


Figure 2: WorldPop Population on a yearly basis from 2000 (top) to 2020 (bottom) and a spatial resolution of 100 m pixels

² <https://hub.worldpop.org/project/categories?id=3>

³ <https://population.un.org/wpp/>

prediction of population density, which is carried out with the Random Forest model. Subsequently, the predicted density layer derived from the RF is employed to redistribute the census counts across a grid, effectively disaggregating the population data to a finer spatial resolution at country level (Stevens et al., 2015). The data is unconstrained, which means that there is no residential settlement mask used to delimit settlement areas and because of that there are no zero allocation of population grid cells.

3.2.2. Land use land cover

As dataset for the Land Use and Land Cover (LULC) within Europe, the CORINE (Coordination of Information on the Environment) programme of the Copernicus Land Monitoring Service (CLMS) implemented by the European Environment Agency (EEA) and the Joint Research Centre (JRC) was used. Corine Land Cover (CLC) is available for the years 1990, 2000, 2006, 2012 and 2018 with a 100 m raster. The high thematic resolution of the dataset allows for the coverage of 44 different LULC classes (Figure 3), which can be grouped into five major categories: artificial surfaces, agricultural land, forest and semi-natural areas, wetlands, and water bodies. The data is generated through the combination of high-resolution satellite imagery with in-situ data, such as topographic maps. The quality control process ensures an accuracy rate exceeding 85%, offering a high level of detail that is invaluable for understanding land cover changes. In this study, the 44 thematic classes are grouped into 11 categories: “vegetation”, “barren”, and “water”. For “built-up” the detailed sub-classes defined by CORINE are used (“Continuous urban fabric”, “Discontinuous urban fabric”, “Industrial or commercial units”, “Road and rail networks and associated land”, “Port areas”, “Airports”, “Construction sites”, “Sport and leisure facilities”).

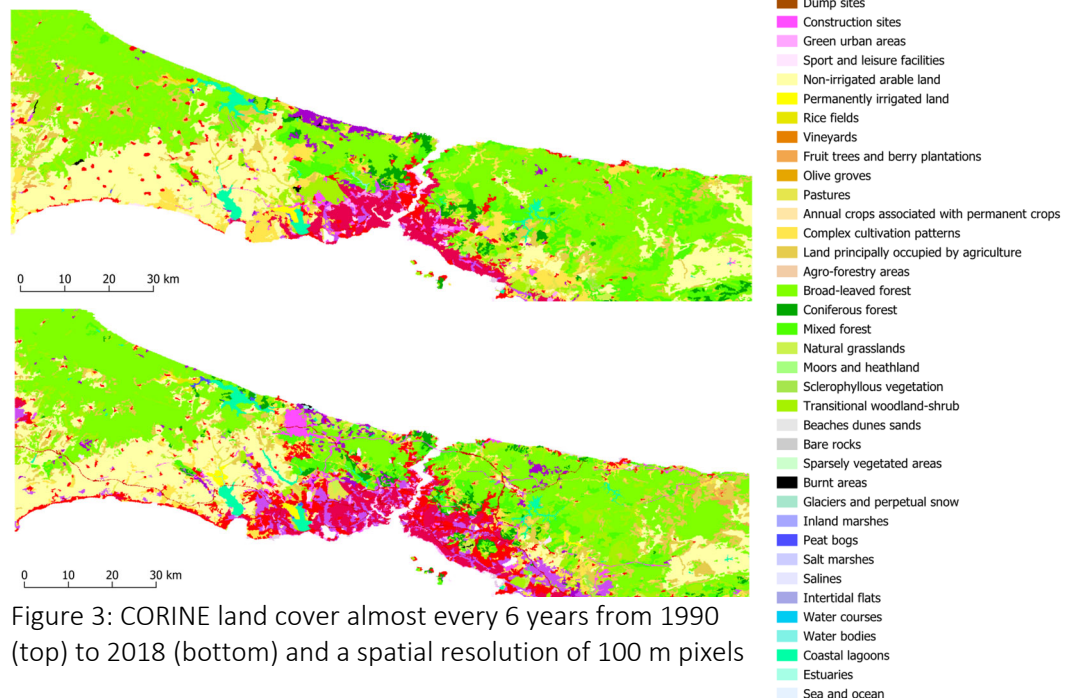


Figure 3: CORINE land cover almost every 6 years from 1990 (top) to 2018 (bottom) and a spatial resolution of 100 m pixels

3.2.3. World Settlement Footprint (WSF) Evolution

An additional gridded settlement dataset is the World Settlement Footprint (WSF) Evolution from the German Aerospace Center (DLR). The dataset outlines the global settlement extent on a yearly basis, beginning in 1985 and continuing until 2015 (Figure 4). It has a spatial resolution of 30 meters. The data is derived from Landsat satellite imagery, with calculations made for the minimum, maximum, mean and standard deviation over time per pixel. These calculations are made for three indices: the normalised difference vegetation index (NDVI), the normalised difference built-up index (NDBI) and the modified normalised difference water index (MNDWI). This is combined with the WSF2015 as a reference for settlement and non-settlement training samples, and the time series is generated with a supervised Random Forest classification (Marconcini et al., 2021). The final product is a raster, with the values indicating the year in which the pixel was identified as a settlement. In order to create a time series, the raster is transformed into a binary representation, with settlement and non-settlement for the available years.

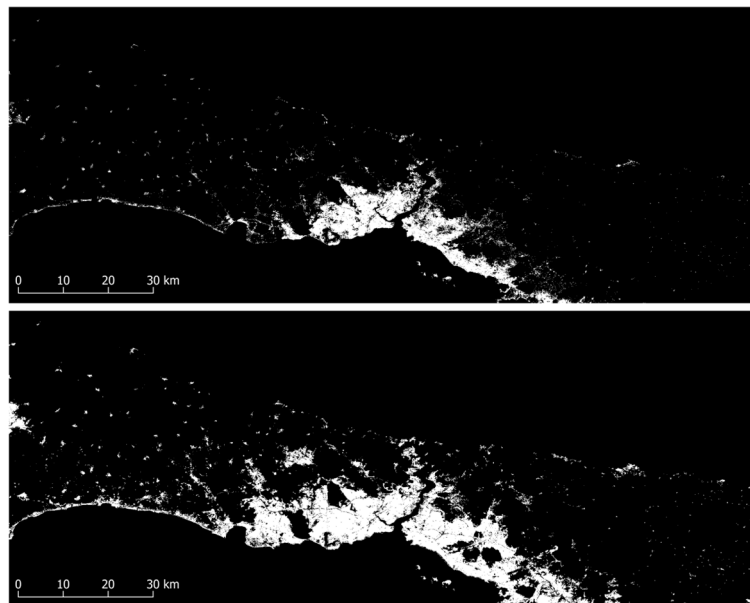


Figure 4: World Settlement Footprint (WSF) Evolution on a yearly basis from 1985 (top) to 2015 (bottom) and a spatial resolution of 30 m pixels

3.3. Driving factors

Additionally, to the time series data we also harvested, harmonized, and aligned to the population grid, a set of geospatial covariates, i.e., driving factors. These variables are relevant for the future prediction of population but also LULC, since they can have an impact on the distribution of both (Zhu, 2022; Geiß et al., 2024). Zhu (2022) provides a comprehensive list of over 50 potential predictor variables that can influence the forecasting of LULC. These have been utilised in previous studies, like Liu et al. (2017) or Gómez et al. (2020). Overall, these variables can be classified into four main categories: environmental, infrastructural, land-use related and social-economic variables. A land-use-related driving factor that varies over time is the

distance to a populated area/urban/built-up. This data was derived from the WorldPop time series by generating a binary raster comprising pixels representing populated and non-populated areas (Figure 5). The Euclidean distance to the nearest populated target pixel was then calculated. Similarly, the same procedure is employed to calculate the distance to water (Figure 6), distance to roads (Figure 8) and distance to city center (Figure 9). The distance to water as environmental variable is derived from the OpenStreetMap (OSM) geodata, which contains information about water bodies, and the water areas in the European Digital Elevation Model (EU-DEM) from the EEA, which has a spatial resolution of 30 m. Another relevant input factor for future modelling of LULC and population is the topography of the terrain, as human settlements primarily appear in flat or moderate slopes (Dobson et al., 2000). Similarly, the European Digital Elevation Model (EU-DEM) is used for this dataset, with the derived slope representing the terrain in the test case (Figure 7). The distance to roads as infrastructural variable is computed by using the road network information provided by OSM. The distance to city center as additional land use related driving factor was calculated by using the middle between the historical city centre and the current central district. With the exception of the distance to urban boundary the driving factors are time-invariant, since variables as terrain or water tend to be relatively static over a longer time spans (Zhu, 2022).



Figure 5: Distance to urban boundary – derived from WorldPop grid with urban and non-urban values and scaled from 0 to 1 from 2000 (l) and 2020 (r)

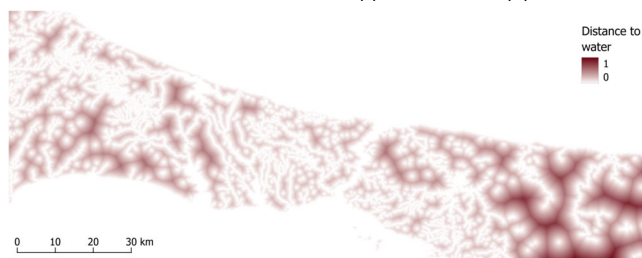


Figure 6: Distance to water – derived from OSM and DEM water surfaces and scaled from 0 to 1

Figure 7: EU-DEM Slope with the slope in degrees and a spatial resolution of 25 m pixels

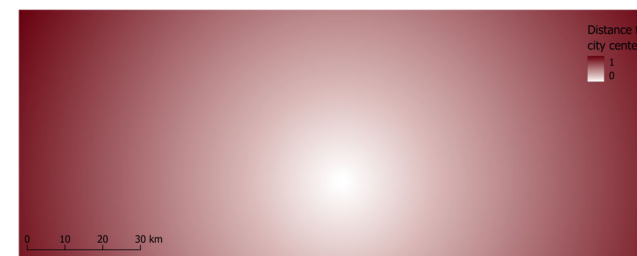


Figure 8: Distance to roads – derived from OSM and scaled from 0 to 1

Figure 9: Distance to city center by using the historical and business city center, scaled from 0 to 1

3.4. Multitask long short-term memory model

As in the approach outlined by Geiß et al. (2024), the time series data comprising the target variables population and LULC are combined with the driving factors as input for the time steps $t_1, t_2, \dots, t_n$. These serve as training data for the single- and multitask long short-term memory (LSTM) models, which can be used for the prediction of future population and LULC. LSTMs represent a version of recurrent neural networks (RNN). RNNs are a type of artificial neural networks (ANN) that can be used to process sequential data (Rumelhart et al., 1986), such as time series or language data. Information is processed with a single element per time but due to the recurrent connections, the information from the previous elements is maintained in the internal state of the hidden units. Consequently, the weights are shared across different time steps (Goodfellow et al., 2016). Due to the “memory” of the network the problem of vanishing and exploding gradients exist, wherefore Hochreiter and Schmidhuber (1997) introduced the concept of long short-term memory (LSTM) networks. An additional feature of LSTMs, compared to RNNs is the incorporation of an LSTM cell, which enables internal recurrence. These cells are equipped with gate and memory cells, which regulate the information flow and can also “forget” used and not needed information. This ability is also learned during the training of the network and enables the LSTM model to learn long-term dependencies (Goodfellow et al., 2016). Due to this advantage LSTMs are useful for the extrapolation of time series data, as demonstrated by Geiß et al. (2024) in their forecasting of future population distribution in Lima.

As previously indicated in the introduction, instead of training related tasks separately/individually within single task learning (STL), Multitask Learning (MTL) can be applied as alternative. Many real-world problems on which models are trained are not simply single-task problems, but rather, they are multitasking problems, since domain-specific information is shared during the training of related tasks. This is employed by MTL as an inductive transfer mechanism with the objective of enhancing the generalisation performance by leveraging this information in a parallel training of these tasks (Caruana, 1997). In comparison to single task learning, multitask learning offers a number of advantages (Vandenhende et al., 2022). For instance, due to the sharing of layers, a reduction in memory footprint is observed. Furthermore, due to less repeated feature calculation they increase inference speeds. When the tasks share complementary information most importantly MTL improves the performance or regulates the specific tasks.

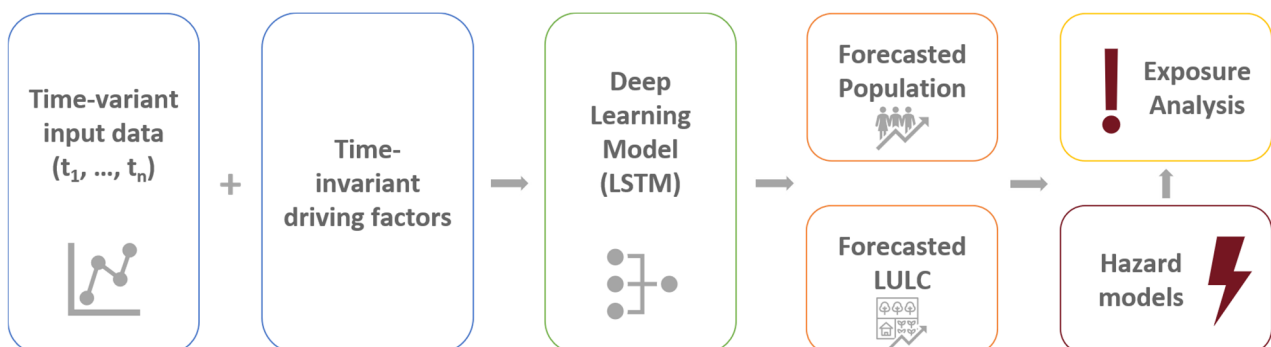


Figure 10: Overview of the current workflow

Within this work the existing method from Geiß et al. (2024), which was based on Zhu (2022), is further developed by adding multitask learning to simultaneously predict future population as one and LULC change as the second task. The LSTM model is used as backbone model, which is shared by both of the two output tasks. To obtain two separate outputs from the shared network, fully connected (linear layers) are added.

4. Experimental Setup

The described datasets are all reprojected and resampled to align with the WorldPop dataset. As coordinate reference system WGS84 (World Geodetic System 1984, EPSG:4326) was chosen, with a 100-meter grid and a spatial dimension of 2385 by 936 pixels for the test case area of Istanbul. The multi-class layer was preprocessed using a one-hot encoding method, resulting in the creation of a separate layer for the final 10 land cover classes. All layers were normalized individually and stacked to create a multidimensional array with the shape (21, 10, 2385, 936). The first dimension shows the time steps from the year 2000 until 2020. The second dimension is the driving factors and also the target variables (pop and LULC), with the aforementioned image size in the last two positions. For the purposes of training and validation, the datasets were subsampled to create smaller, overlapping images of 256 by 256 pixels. Similar to Geiß et al. (2024), not all the time steps are used, but a three-year interval. This represents an optimal balance between capturing the timely change and forecasting capability, while also ensuring sufficient input time steps to maintain efficient training. As illustrated in Figure 11 six time steps were selected for training purposes (2000, 2003, 2006, 2009, 2012, 2015), and an identical number was employed for validation (2003, 2006, 2009, 2012, 2015, 2018). For both the final time step is used as target year for the validation, while the initial time steps are used for training the model.

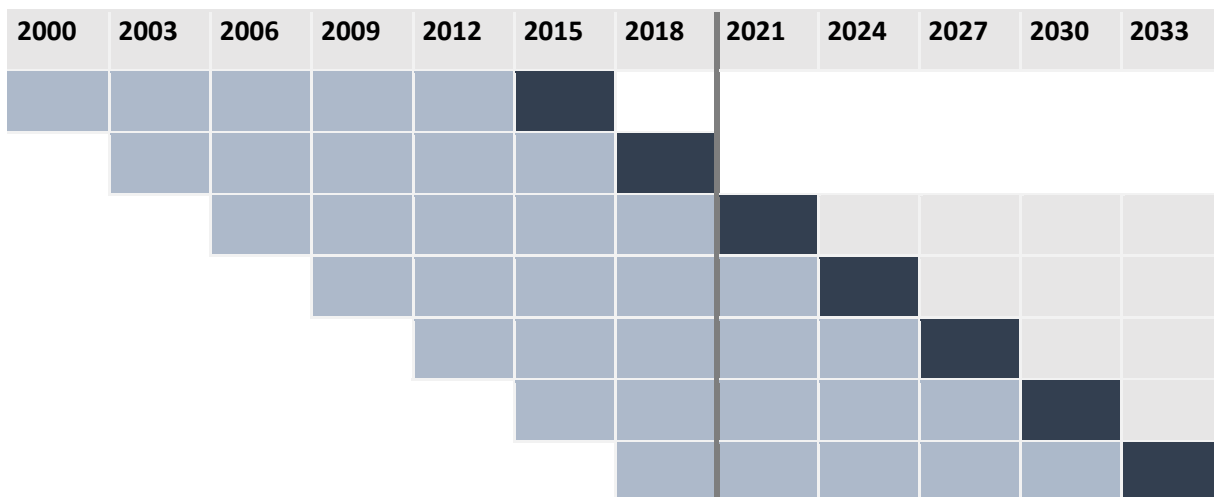


Figure 11: Training, validation and forecasting time steps for a 3-year interval

5. Experimental results and outlook

With the trained model, future time steps of population distribution, alongside land use and land cover (LULC) changes, can be forecasted. This model employs a sliding time window approach, allowing forecasts over multiple future intervals. The maximum forecastable time steps are determined by the chosen interval (e.g., annually or triennially) and the extent of available historical data. Previously forecasted years are iteratively used as inputs to project further time steps. For instance, in the case of Istanbul, population distribution forecasts with a three-year interval can be extended to 2033. The first experimental output of population forecasting (Figure 12, 13 & 14) shows the potential of the method and implicates next steps for evaluation measures and further statistical analysis. The model was validated using 2018 as a reference year, with ground-truth data providing preliminary accuracy metrics. The population prediction achieved a Root Mean Squared Error (RMSE) of 0.2 (Hodson, 2022), while LULC classification demonstrated an overall accuracy of 90% (Parashar, 2023).

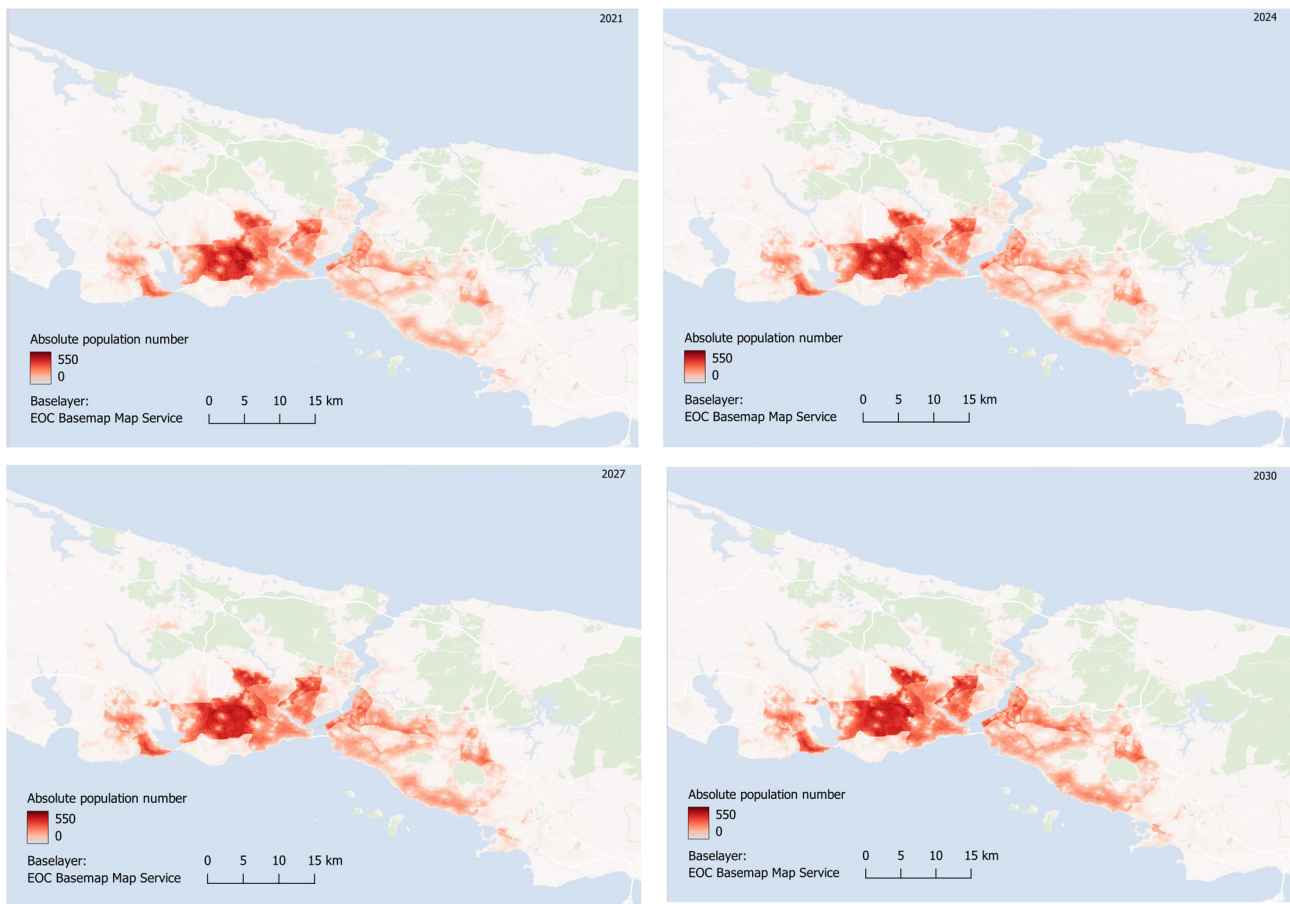


Figure 12: Forecasted time steps (2021, 2024, 2027, 2030) that show possible future population distribution based on the past trend (preliminary result)

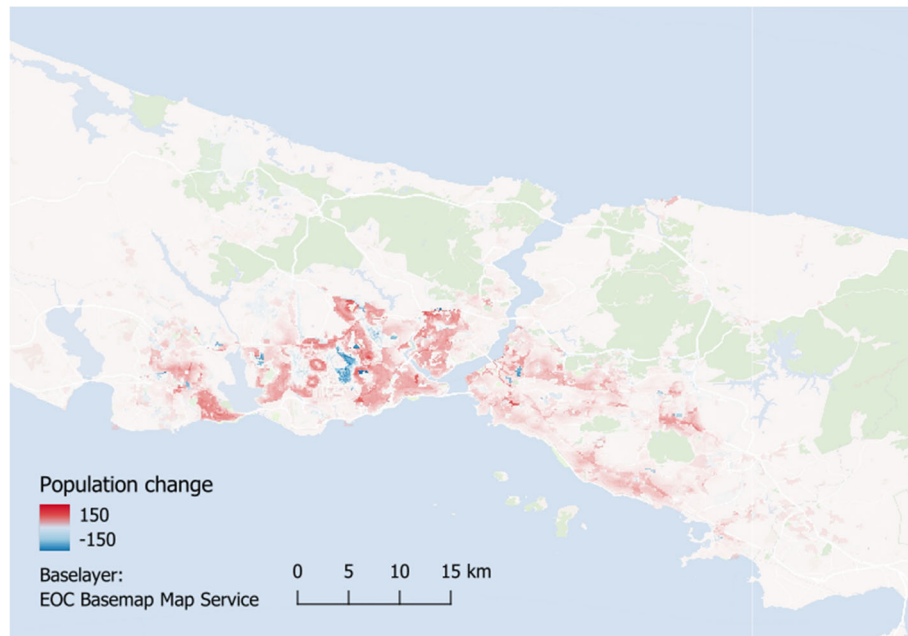


Figure 13: By comparing the forecasted population distribution for 2030 with the WorldPop values from 2018, areas with population increase and decrease can be identified (preliminary result in absolute numbers).

The resulting spatial projections of population distribution in Istanbul provide essential data for multiple applications. One key use is to analyse areas of population growth and decline by comparing projections against the 2018 WorldPop dataset (Figure 13). This information on anticipated population shifts informs models of natural hazard exposure, allowing predictions of potential future impacts from earthquakes, wildfires, floods, and other events. In the case of Istanbul, the forecasted exposure datasets can be integrated with earthquake models and other relevant data to generate hazard risk maps.

Based on the results obtained, potential next steps are outlined in the following. A further step in the evaluation of the approach is to compare it with statistical population estimations based on demographic information, including data on birth and death rates and migration movements. A comparison may be made at the city, district, neighbourhood or other level, depending on the availability of data. The migration or potential relocation of activities in the past is thereby represented in a geographic context, with changes in population numbers and distribution mapped using the World Pop data set.

As the method is a trend analysis, external variable factors such as a pandemic or a political decision are not included in the training process. Next steps could be to consider how this information can be spatially coded and included in the model.

The subsequent stages of the process, in addition to the consideration of the data required for WP2, may involve the aggregation of population change at the district or zoning level for subsequent integration into WP2. In the case of data availability, the aggregated data in conjunction with zoning plans of Istanbul can also be employed for future planning purposes.

By further developing the methodology within the PARATUS project, the described pathway to predict population is followed in a more exhaustive way. However, PARATUS innovatively aims to provide a complete picture of risk by modelling and anticipating the exposure and vulnerability of the built environment in addition to population forecasts. LULC projections serve as foundational data for mapping future changes in built environments, which can be further refined by incorporating detailed geospatial datasets. As an example, an overview of LULC prediction for 2030 is mapped in figure 14.

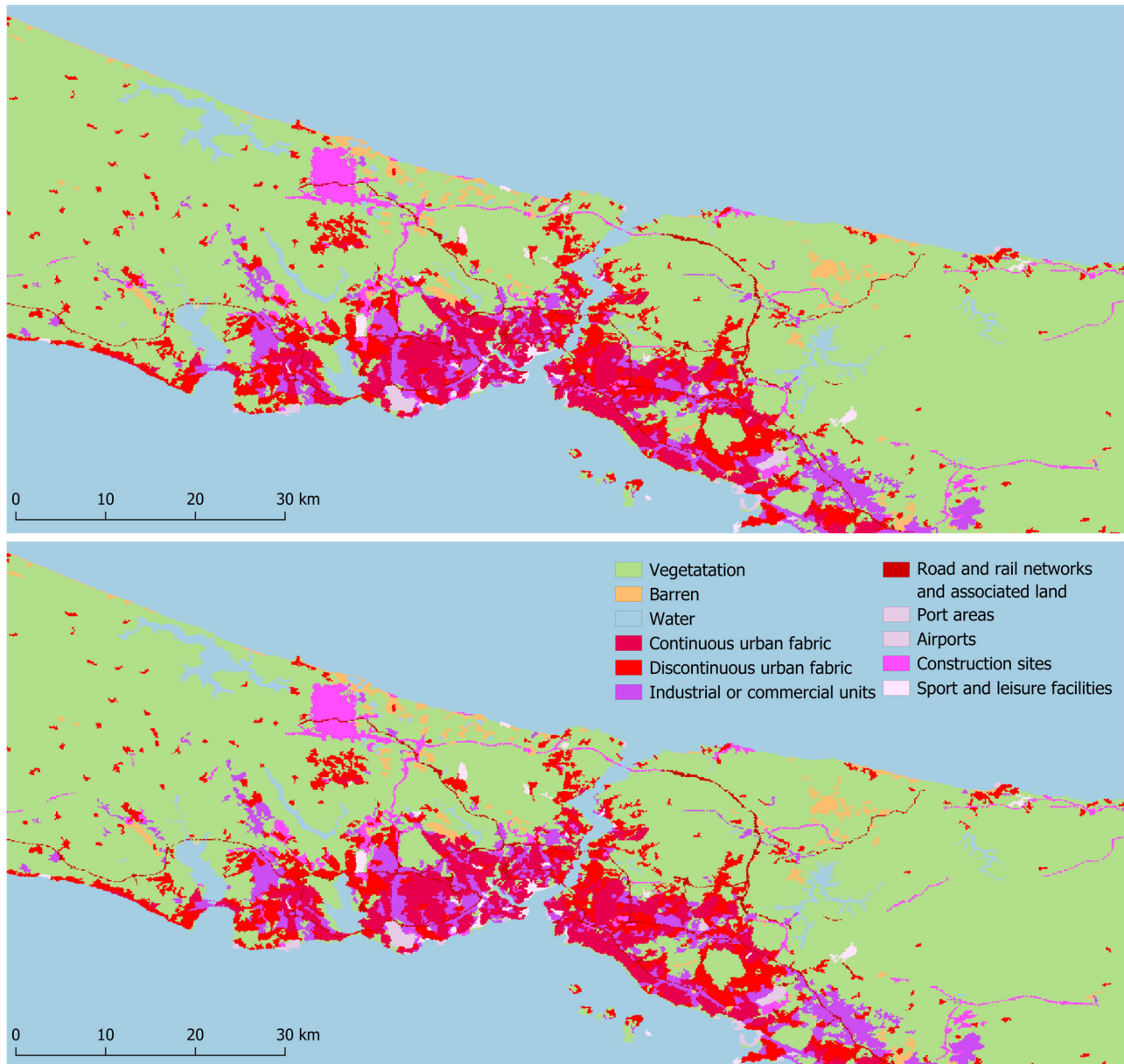


Figure 14: Prediction of the LULC change for 2030 (bottom) compared to the CORINE LULC from 2018 (top)

Additionally, to the obtained results, a possible next step can be a statistical representation of the classes on an aggregation level, that is useful for further analysis in other tasks, for instance combined with before mentioned zoning plans of Istanbul to better show land use changes.

The output of this deliverable is input for the WP2 “Today’s future: how current hazard and risk may change”. WP2 focuses on the development of a generic methodology for a systemic multi-sectoral and multi-hazard risk assessment and its application to the four application case studies using the outputs of WP1 and the available information from earlier European projects and Copernicus services. The produced risk components for systemic risk assessment from this deliverable 1.5 are therefore valuable input data for WP2.

As a future outlook for potential hazard risk maps, a comparable methodology was employed in Lima and Callao, Peru, where population growth was projected at three-year intervals up to 2035. When integrated with simulations of seismic and tsunami hazards, these projections (Figure 15) demonstrated that population growth is anticipated in regions characterised by high peak ground acceleration as well as tsunami inundation areas (Geiß et al., 2024). For instance, this can also be done in a similar way for the city of Istanbul in the context of a possible earthquake scenario.

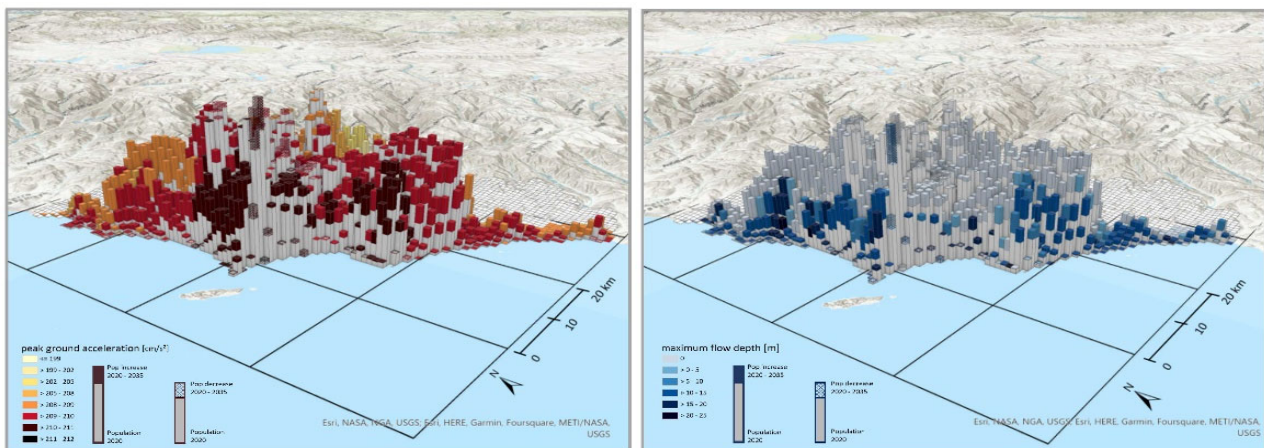


Figure 15: Maps of the predicted population of Lima affected by earthquakes and tsunamis for the year 2035, alongside corresponding hazard intensities, illustrate the changes compared to 2020. In these maps, solid grey bars represent the population data from 2020, while additional coloured or textured bars indicate the estimated population changes—either increases or decreases—by 2035. The colour coding on these bars correlates with the intensity of the respective hazards (Geiß et al., 2024).

6. References

- Aravena Pelizari, P., Geiß, C., Groth, S., Taubenböck, H., 2023. Deep multitask learning with label interdependency distillation for multicriteria street-level image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 204, 275–290.
- Caruana, R., 1997. Multitask Learning. *Machine Learning* 28, 41–75.
- Chen, Y., Li, X., Huang, K., Luo, M., Gao, M., 2020. High-Resolution Gridded Population Projections for China Under the Shared Socioeconomic Pathways. *Earth's Future* 8, e2020EF001491.
- Clarke, K., 2014. Cellular Automata and Agent-Based Models, in: *Handbook Regional Science*. pp. 1217–1233.
- Clarke, K.C., 2008. Mapping and Modelling Land Use Change: an Application of the SLEUTH Model, in: Pettit, C., Cartwright, W., Bishop, I., Lowell, K., Pullar, D., Duncan, D. (Eds.), *Springer, Berlin, Heidelberg*, pp. 353–366.
- Cremen, G., Galasso, C., McCloskey, J., 2022. Modelling and quantifying tomorrow's risks from natural hazards. *Science of The Total Environment* 817, 152552.
- Gagniuc, P.A., 2017. *Markov Chains: From Theory to Implementation and Experimentation*. Wiley.
- Geiß, C., Maier, J., So, E., Schoepfer, E., Harig, S., Gómez Zapata, J.C., Zhu, Y., 2024. Anticipating a risky future: long short-term memory (LSTM) models for spatiotemporal extrapolation of population data in areas prone to earthquakes and tsunamis in Lima, Peru. *Natural Hazards and Earth System Sciences* 24, 1051–1064.
- Gómez, J.A., Patiño, J.E., Duque, J.C., Passos, S., 2020. Spatiotemporal Modeling of Urban Growth Using Machine Learning. *Remote Sensing* 12, 109.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press.
- Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Computation* 9, 1735–1780.
- Hodson, T.O., 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development* 15, 5481–5487.
- Islam, M., VS, V., Lim, C.M., Ren, H., 2021. ST-MTL: Spatio-Temporal Multitask Learning Model to Predict Scanpath While Tracking Instruments in Robotic Surgery. *arXiv*.
- Jin, N., Wu, J., Ma, X., Yan, K., Mo, Y., 2020. Multi-Task Learning Model Based on Multi-Scale CNN and LSTM for Sentiment Classification. *IEEE Access* 8, 77060–77072.
- Karimzadeh, M., Schwegler, S.M., Zhao, Z., Braun, T., Sargento, S., 2021. MTL-LSTM: Multi-Task Learning-based LSTM for Urban Traffic Flow Forecasting, in: *2021InternationalWirelessCommunications Mobile Computing (IWCMC)*. pp. 564–569.
- Krüger, C., 2016. *Uncertainties in Land Change Modeling*.
- Kubota, Y., Ohira, Y., Shimizu, T., 2022. Attention-based Contextual Multi-View Graph Convolutional Networks for Short-term Population Prediction.
- Kundak, S., Goksu, C., Arslanli, K.Y., Asici, A.A., Kalkanlı, D., 2024. Spatial and Temporal Impact Chain Analysis on Seismic Risks of Istanbul, in: Erberik, M.A., Askan, A., Kockar, M.K. (Eds.), *Proceedings 7th International ConferenceEarthquakeEngineering Seismology*. Springer Nature Switzerland, Cham, pp. 415–423.
- Liu, X., Liang, X., Li, X., Xu, X., Ou, J., Chen, Y., Li, S., Wang, S., Pei, F., 2017. A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning* 168, 94–116.
- Maithani, S., 2009. A neural network based urban growth model of an Indian city. *Journal of the Indian Society of Remote Sensing* 37, 363–376.
- Marconcini, M., Metz- Marconcini, A., Esch, T., Gorelick, N., 2021. Understanding Current Trends in Global Urbanisation - The World Settlement Footprint Suite, in: *GI_Forum*. pp. 33–38.

- United Nations (UN), 2024. World Population Prospects 2024: Summary of Results (No. UN DESA/POP/2024/TR/NO. 9). United Nations Department of Economic and Social Affairs, Population Division, New York.
- Pan, X., Liu, Z., He, C., Huang, Q., 2022. Modeling urban expansion by integrating a convolutional neural network and a recurrent neural network. *International Journal of Applied Earth Observation and Geoinformation* 112, 102977.
- Parashar, N., 2023. What is an Accuracy Score and How to Check it?
- Parsons, T., 2004. Recalculated probability of $M \geq 7$ earthquakes beneath the Sea of Marmara, Turkey. *Journal of Geophysical Research: Solid Earth* 109.
- Portoghese, L., 2024. Case Studies. PARATUS.
- Qian, Y., Xing, W., Guan, X., Yang, T., Wu, H., 2020. Coupling cellular automata with area partitioning and spatiotemporal convolution for dynamic land use change simulation. *Science of The Total Environment* 722, 137738.
- Ritchie, H., Rosado, P., Roser, M., 2024a. Is the number of natural disasters increasing? Our World in Data.
- Ritchie, H., Rosado, P., Roser, M., 2024b. Natural Disasters. Our World in Data.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLOS ONE* 10, e0107042.
- Vandenhende, S., Georgoulis, S., Van Gansbeke, W., Proesmans, M., Dai, D., Van Gool, L., 2022. Multi-Task Learning for Dense Prediction Tasks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 3614–3633.
- Yan, Y., Li, G., Li, Q., Zhu, J., 2024. Enhancing Hydrological Variable Prediction through Multitask LSTM Models. *Water* 16, 2156.
- Zhang, Y., Yang, Q., 2022. A Survey on Multi-Task Learning. *IEEE Transactions on Knowledge and Data Engineering* 34, 5586–5609.
- Zheng, Z., Zhang, G., 2020. The Prediction of Finely-Grained Spatiotemporal Relative Human Population Density Distributions in China. *IEEE Access* 8, 181534–181546.
- Zhu, Y., 2022. A Deep Learning Framework for Investigating Spatio-temporal Evolution of Land Use and Land Cover Patterns.
- Zhu, Y., Geiß, C., So, E., Bardhan, R., Taubenböck, H., Jin, Y., 2024. Urban expansion simulation with an explainable ensemble deep learning framework. *Heliyon* 10, e28318.