FORECASTING URBAN FUTURES: EVALUATING GLOBAL LAND USE DATA SENSITIVITY FOR REGIONAL GROWTH SIMULATION IN THE RUHR METROPOLITAN AREA

Andreas Rienow

With 8 figures and 5 tables
Received 24 January 2024 · Accepted 11 April 2024

Summary: In recent years, numerous multitemporal global land use and land cover products have been published acting as valuable source for training spatially explicit geosimulation models forecasting urban growth. However, there is a notable gap in research that specifically addresses the sensitivity of models traing with those data sets when it comes to regional modeling purposes. Accordingly, the objectives of this study were to calibrate, validate, and employ global urban input datasets for the regional simulation of urban growth by the year 2030. The SLEUTH urban growth model, focused on the metropolitan area of the Ruhr, Germany, was calibrated using the Global Human Settlement Layer, World Settlement Footprint Evolution, historical OpenStreetMap data, and a Digital Land Cover Model for Germany. The goal was to compare the results in terms of accuracy, certainty, quantity, and allocation, particularly in urban areas susceptible to floods and heat. While all models achieved high accuracy levels concerning quantity and allocation, the extent of new settlements varied from 40.77 km2 to 477.91 km². The models based on World Settlement Footprint and OpenStreetMap exhibited higher certainty and lower stochasticity. As the simulated urban growth increased, there was a corresponding rise in the likelihood of allocating new settlements in areas affected by natural hazards. While all models presented a similar relative portion of new settlement areas impacted by floods, variations emerged in terms of areas affected by unfavorable thermal conditions. This study underscored the potential use of historical OpenStreetMap data in training cellular automation for geosimulating future settlement growth. Furthermore, it highlighted the applicability of global Earth observation-based urban datasets for regional geosimulation and explored the impacts of diverse input data on the accuracy, certainty, quantity, and allocation performances in simulating future conditions.

Keywords: Urban development, SLEUTH, OpenStreetMap, Ruhr Metropolitan Area, Global Human Settlement Layer, World Settlement Footprint

1 Introduction

By 2050, two-thirds of the population is expected to live in cities. With this continuously increasing urban population and its footprint, sustainable urban development requires the assessment, mapping, and modelling of urban environments with high spatial detail. This is particularly true since the impacts of progressing anthropogenically induced climate change also affect an increasing number of cities and agglomerations in temperate climate zones (Adamo et al. 2012, Else 2021, Oleson et al. 2015, Rienow et al. 2022). Machine learning-based urban models are recognized as one of the main tools for urban monitoring in terms of description, explanation, planning, and future prospects of urban growth on disaster risk management (Chien et al. 2020, Hassan & Elhassan 2020, Krelaus et al. 2021, Wang & Upreti 2019). These models rely on accessible, accurate, and temporally and spatially detailed input data on settlement patterns. Recently, satellite-based global urban datasets have emerged, providing researchers with human settlement footprints with associated historical development, such as the Global Human Settlement Layer (GHSL) (Corbane et al. 2018a, 2019) and the World Settlement Footprint Evolution (WSF Evo) (Marconcini et al. 2021, 2020). Furthermore, the Ohsome Application Programming Interface (API) (Raifer et al. 2019) enables scientists to access historic OpenStreetMap (OSM) data, making it the most used volunteered geographic information (VGI) system worldwide (Jokar Arsanjani et al. 2015). Several studies have used OSM information to derive land use and land cover (LULC) products. Estima & Painho (2015) investigated the potential of OSM data for LULC maps production, using the CORINE Land Cover inventory as a reference. Schultz et al. (2017) filled the OSM data gaps using available open source remote sensing data. In addition, Patriarca et al. (2019) developed an automated conversion of OSM data into LULC maps. The transition to raster data format and the implementation of procedural com-

https://doi.org/10.3112/erdkunde.2024.01.04 ISSN 2702-5985 (Online)

ponents using OSM data led to notable performance improvements, with no notable positional distortions that would impair the usability of the final outcome in subsequent case scenarios. The LandSense project hosts the WebService, providing LULC information derived from OpenStreetMap with global coverage (https://osmlanduse.org). Zhou et al. (2015) applied cellular automaton (CA) with a global urban LULC input dataset, derived from the commercial product LandScan™, to compare global urban LULC products and their applicability for regional geosimulation studies. The regional scale is often deemed optimal for spatially explicit modeling due to its ability to capture detailed local variations while maintaining a broader perspective. This scale offers a balance between granularity and comprehensiveness, making it suitable for various analyses such as land use planning, environmental impact assessment, and hazard monitoring. On the other hand, global settlement products provide a comprehensive view of human settlements worldwide, offering valuable insights into global trends and patterns. However, they may lack the fine-scale detail required for local decision-making and can be prone to inaccuracies, particularly in areas with limited data availability or complex terrain. Therefore, while global settlement products are valuable for broad-scale analyses, they may not always suffice for detailed, localized studies (Liu et al. 2020b). Pesaresi & Politis (2023) utilized spatial-temporal interpolation and extrapolation for predicting global settlement patterns in 2025 and 2030 based on the GHSL with a resolution of 100 m. They applied a rank-optimal spatial allocation method to resolve a built-up prediction by combining static and dynamical components based on empirical associations between specific land form combinations and human settlement development gleaned from remotely sensed data. With regard to WSF Evo, Wang et al. (2022b) conducted a regional case study for comparing the spatio-temporal matrix (STM) approach and SLEUTH Urban Growth Model (UGM) for the prediction of future settlement growth for the urban agglomerations of Surat (India), Ho-Chi-Minh City (Vietnam), and Abidjan (Ivory Coast). They conclude that STM-based models outperform SLEUTH-UGM but are not able to predict spontaneous settlement growth. Furthermore, they state that the WSF Evo product restricts the modelling and evaluation of the model outputs with real measurements to certain constellations due to its availability from 1985-2015. Additionally, studies like the aforementioned do not focusing on the sensitivity of global LULC products on spatially explicit geosimulation models and how the input data might change the outcome especially when it comes to environmental impacts and hazard monitoring. This is also true for studies comparing the geosimulation potential of open global LULC products and their applicability at the regional level.

Therefore, the objective of this study was to carry out a regional study on future settlement developments in the context of environmental hazards like floods and heats stress. This was executed through the calibration, validation, implementation of WSF, GHSL, and OSM data for simulation of urban growth. Global Earth observation-based settlement datasets were utilized for regional planning purposes, which had vast temporal extensions (1975–2015 GHSL), the highest temporal resolution (annually from to 1985–2015, WSF), and most up-to-date data (OSM). Additionally, historic OSM data were applied for the implementation of an urban CA as a common urban growth model. The CA used in this study was the SLEUTH Urban Growth Model (Clarke et al. 1997a), which is one of the most recognized urban growth models in the world with low requirements regarding the cellular urban input data (Clarke & Johnson 2020, Jantz et al. 2010, Liu et al. 2020b, Rienow et al. 2014, 2015, Rienow & Stenger 2014, Saxena et al. 2021, Wang et al. 2022a, Zhou et al. 2019). The study area is the metropolitan area of Ruhr in Western Germany and was chosen for its challenging polycentric structure for bottom-up machine learning models (Rienow et al. 2014, Rienow & Stenger 2014, Wang et al. 2022a). The three global datasets were compared with a national-administrative dataset, assuming a higher accuracy of settlement information than the global datasets. The Digital Land Cover Model (DLM) for Germany (German "Digitales Landbedeckungsmodell für Deutschland", LBM-DE, until 2012 DLM-DE) was based on selected areal object types of the authoritative ATKIS® Basis-DLM including areas of settlement, traffic, vegetation, and water bodies, which were adapted as a modified form to tailor for the specific requirements of CORINE Land Cover nomenclature. The minimum mapping area of the dataset was 1 ha (Hovenbitzer et al. 2014). The spatial simulation results for 2030 were combined with two hazard maps: frequent flood risk and areas with a potential for heat stress as proxy for heat stress (Siedentop et al. 2014). Hence, the subsequent research questions guided this study:

- How accurate and certain can the SLEUTH "Urban Growth Model reduced" (UGMr) data geographically simulate regional urban growth in a polycentric region based on global settlement datasets, in comparison to that based on a regional LULC product?

- How do the CA results differ in quantity and allocation in urban environments affected by floods and heats when calibrated with different OSM mapping periods?
- How do the CA results, based on WSF and GHSL, differ in quantity and allocation in urban environments affected by floods and heats?

This study is structured as follows: Section 2 introduces the study area; Section 3 describes the data and methods, which includes the CA SLEUTH Urban Growth Model, the utilized global settlement products and their harmonization, as well as further input data and geodata on environmental hazards. Section 4 presents the results of historic urban development and the calibration, validation, and settlement simulation of 2030. Section 5 discusses the limitations of the applied methods, and Section 6 states the conclusions of the study.

2 Study area

The study area lies in North Rhine-Westphalia (NRW) in western Germany (Fig. 1). It extends from the Lower Rhine Basin in the west to the Westphalian Plane in the north and the Rhenish Massif in the south. With its polycentric and administratively fragmented structure, but homogenous and extensive urban area, Ruhr is a unique urban entity. Eleven cities and four districts form the largest agglomeration (1,152 people per km^2) in Germany, and at 4,435 km^2, it is the fifth largest urban region in Europe. The largest cities of Ruhr are, in descending order, Dortmund, Essen, Duisburg, and Bochum, with populations varying between 364,000 (Bochum) and 592,000 (Dortmund).

The rapid economic, demographic, and morphogenetic growth of Ruhr at the turn of the 20th century was owing to construction during the age of industrial mining, spread of ironworks, railway expansion, and the foundation of Zollverein in the mid-nineteenth century. This era of prosperity continued until after World War II. The cultural identity of the people of Ruhr is closely intertwined with industrial heritage. This is reflected in the way the elements of the natural landscape have been utilized; the tectonic faults provided carbon and ore that were refined in the steel works, and the rural hinterland provided agricultural products. There are three rivers which cross the Ruhr region, from the mountainous east to the western Rhine valley, which also provide different social functions. The Ruhr River acted as a potable water resource, the Lippe River was utilized as a cooling water pool, and wastewater was diverted into the infamous Emscher River. The Rhine River has always served as an international trade route, and the Duisburg harbor has become the world's largest inland port. Today, numerous renaturation sanctions, university buildings, and urban entertainment centers turn the image of the Ruhr as a sealed moloch with nights aflamed by smelting furnaces and freshly washed clothes colored black by smut into a cliché. Only the fine-meshed net of autobahns, museum constructions, and multicultural population provide evidence of the region's industrial past (Hospers & Wetterau 2018).

3 Data and methods

3.1 SLEUTH urban growth model reduced

The cellular automaton SLEUTH was developed for modeling changes in land-use (Clarke et al. 1997a, Clarke 2008). It was originally conceived for more than one land-use class. Most studies in the scientific literature identify SLEUTH with one of its two sub-models. The first is the Clarke Urban Growth Model (UGM) (Clarke et al. 1997b), which focuses exclusively on simulating urban growth. The second SLEUTH component is the Deltatron Land Cover Model, which was not used in this study.

For many applications, the amount of data required to calibrate the UGM is difficult to acquire. It was therefore reprogrammed by Goetzke (2012) and implemented in the eXtendable Unified Land Use Modeling Platform (XULU©), a JAVA-based modeling environment developed at the University of Bonn (Goetzke 2012, Judex 2008, Schmitz et al. 2007).

The standard calibration evaluation method of the UGM has been replaced by multiple resolution validation (MRV) (Pontius et al. 2004, Pontius & Malizia 2004, Rienow & Goetzke 2015), which compares a simulated map with an observed map at different spatial resolutions. High resolutions had a higher weighting than low resolutions. Generally, MRV is to attenuate the impact of localization errors by extending the conventional cell-by-cell comparison. In addition, it considers the similarity of the entire neighborhood of a cell. Thus, the fuzziness of simulated maps has been taken into account (Costanza & Maxwell 1991, Visser 2004)

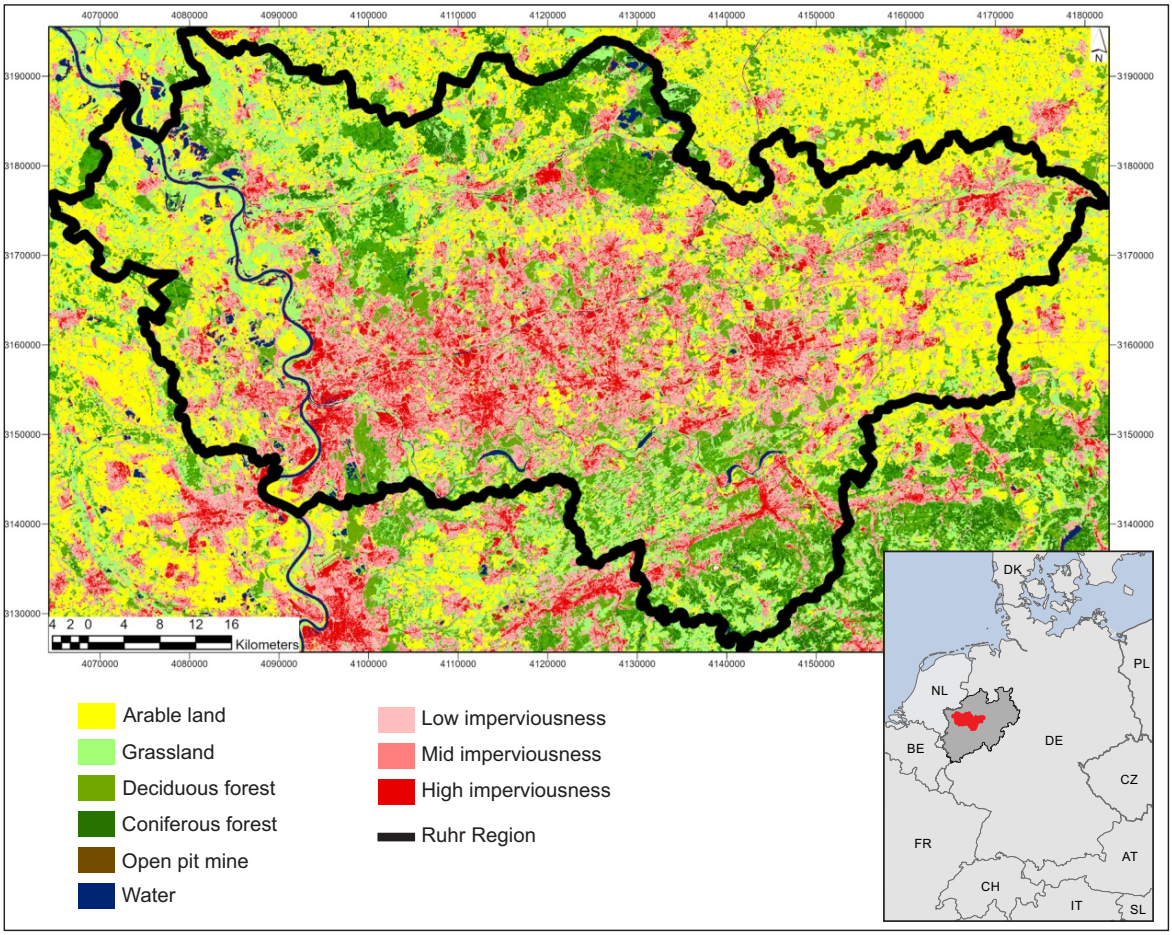

Fig. 1: Study Area: The Ruhr Metropolitan Area

by enabling the accurate simulation of spatial patterns by correctly classifying in a defined neighborhood. Additionally, urban land use calibration with MRV requires only two maps, for the start and end years of calibration. The modified version of the UGM is referred to as the "Urban Growth Model reduced" (UGMr).

The base data of the UGMr includes the settlement pattern of urban land use, classified as either urban or non-urban. The slope and a map of transportation infrastructure are mandatory for the model algorithm. An exclusion layer is used to incorporate political constraints, such as natural reserves or areas that are excluded for urban growth, such as water bodies (0 = growth possible, 1 = growth not possible); this layer is an optional component for the model.

One growth cycle of the UGMr represents one year of urban growth. Five growth coefficients (dispersion, breed, spread, slope, and road gravity) define the four growth rules of the UGMr subsequently performed for each growth cycle. (Fig. 2): (i) The first rule is spontaneous growth, representing the random emergence of new urban areas. This is determined using the dispersion coefficient. (ii) Breed growth refers to newly urbanized cells that can act as core areas for urbanization in direct neighborhoods. (iii) Edge growth represents radial urban sprawl and the infill of existing urban areas. This is regulated by the spread coefficient. The UGMr simulates extensive edge growth in a Moore neighborhood of at least three urbanized cells. (iv) The fourth growth rule is road-influenced growth. Starting from a cell urbanized during the current growth cycle, the next road in a certain neighborhood is selected, and a temporary cell is relocated along the road to its final position, which is influenced by the dispersion coefficient. Each probable new urban cell selected by a growth rule was tested against the local slope and exclusion information before it was urbanized.

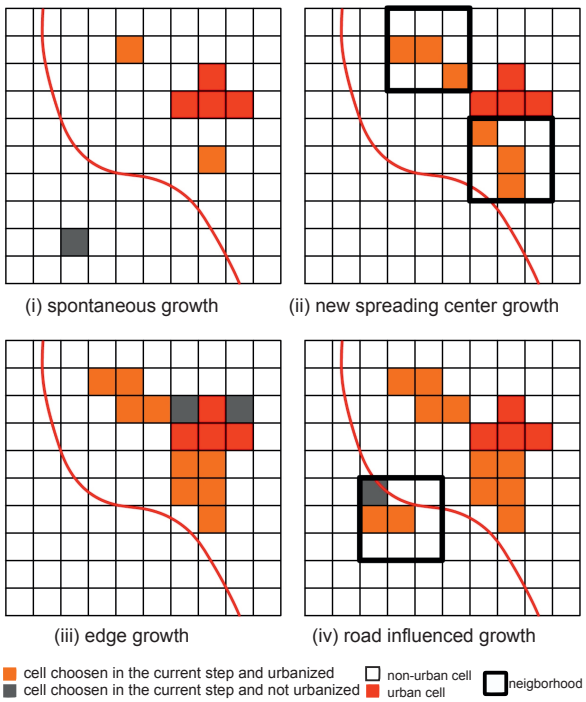

Fig. 2: Modeling steps in the SLEUTH model (Rienow & Goetzke 2015)

An urban CA is a typical bottom-up approach that attempts to capture the self-organizing nature of complex urban systems (Batty 2008, Benenson & Torrens 2004). No ex-ante knowledge of the parameters are required and they were detected by inductive reasoning during model calibration. Hence, the calibration process is based on the brute-force method. Each combination of the particular growth coefficients between 0 and 100 was tested. An assessment of all possible parameter combinations can be very time-consuming; therefore, the calibration procedure was performed in several steps, starting with a coarse evaluation, and then refining the results at several intervals (Goetzke 2012, Rafiee et al. 2009, Wu 1998). Subsequently, the combination that achieved the best MRV results was chosen. For the simulation run, 100 Monte Carlo (MC) iterations were used to depress the model's stochastic nature. The result of the MC iterations was a raster showing, for each cell, how often the model selected it for urbanization (Badmos et al. 2019, Rienow & Goetzke 2015).

In this study, the SLEUTH UGMr was calibrated and validated using different datasets on settlement patterns to compare the results in terms of accuracy, quantity, and allocation behavior.

3.2 SLEUTH input data

3.2.1 Settlement datasets

Open Street Map

OSM data is one of the most widely used VGI systems worldwide (Jokar Arsanjani et al. 2015). Geodata are stored in a topological data structure based on four data primitives. These are nodes (points with geographic positions); ways (ordered lists of nodes representing polylines and polygons); relations (ordered lists of nodes and ways representing relationships among nodes and ways), and tags. The tags were stored as key-value pairs with a recommended ontology (https://wiki.openstreetmap.org/wiki/Map_features). To access historic OSM data, the Heidelberg Institute for Geoinformation Technology developed the Ohsome API as an OSM History Analyzer (Raifer et al. 2019). The primary focus of the Ohsome API is intrinsic quality assessment. An additional benefit is the possibility of analyzing the development of residential areas and certain amenities. Figure 3 presents the historic OSM mapping activities on nodes, ways, and relations in Ruhr from 2008 to 2021, for the key land use.

A Python script was developed for Jupyter Notebook accessing the Ohsome API and included relevant OSM tags to map the development of settlement land use types. Table 1 presents the features utilized for the key "land use" in mapping settlement uses (Fonte et al. 2019, Patriarca et al. 2019, Schultz et al. 2017). Reliability and consistency were considered using the OSM data, commencing in 2012.

World Settlement Footprint evolution

The WSF evolution is a 30 m resolution dataset outlining the global settlement extent on a yearly basis from 1985 to 2015. The processing method is described in Marconcini et al. (2021).

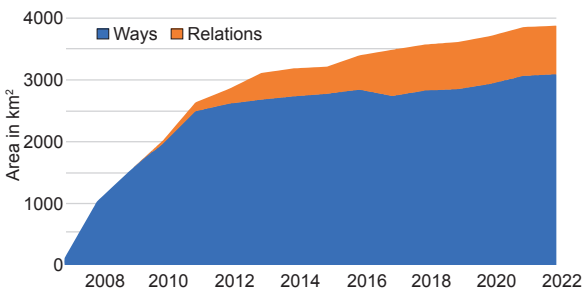

Fig. 3: Effect of mapping period: Ohsome feature counts for the key land use in the Ruhr region

Based on the assumption that settlement growth has occurred over time, all pixels catergorized as non-settlements in the WSF representing the year 2015 (WSF2015) (Marconcini et al. 2020) were excluded from the analysis. For each past target year, all available Landsat-5/7 scenes were gathered and key temporal statistics for the different spectral indices were extracted. Working backwards from 2015, the settlement and non-settlement training samples for the given target year were iteratively calculated. Finally, a boolean Random Forest classification was performed. The WSF evolution datasets are organized into 5,138 GeoTIFF files, each of which refers to a portion of a 2 × 2 degree size, which equates to approximately 222 × 222 km on the ground. An accuracy assessment can be found in Marconcini et al. (2021).

Global Human Settlement Layer built-up area grid

The GHSL built-up area grid (GHS_BUILT_LDSMT_GLOBE_R2018A) was based on Landsat and was developed for 1975, 1990, 2000, and 2014 (Corbane et al. 2019 & 2018b, EC JRC et al. 2019). The product was developed and disseminated by a team at the European Commission Joint Research Center (JRC) and was based on 33,202 images (Florczyk et al. 2018). The main product is the multitemporal classification layer of built-up areas derived from the Global Land Survey (GLS) Landsat image collections (GLS1975, GLS1990, GLS2000, and ad-hoc Landsat 8 collection of 2013/2014). It builds on the Symbolic Machine learning method designed for remote sensing big data analytics (Pesaresi et al. 2016). Corbane et al. (2019) has provided a detailed explanation of the processing steps and usage of the GHSL Sentinel-1 dataset (GHS_BUILT_S1NODSM_GLOBE_R2018A) as a learning dataset, as well as the multitemporal accuracy assessment. Recently, the JRC released an update to the GHSL dataset (EC JRC 2023, Pesaresi & Politis 2023). This is the GHS built-up surface grid and was derived from Sentinel-2 composite and Landsat data, and multitemporal data from 1975 modeled to 2030 (GHS-BUILT-S R2023A). All years are at a resolution of 100 m, barring 2018, which is at 10 m. In addition, it continuously stores built-up surfaces continuously and non-boolean. It also focuses on residential areas for population disaggregation purposes. The product (GHS-BUILT-S R2023A) comprises of boolean change maps on built-up areas in general.

Digital Land Cover Model for Germany

The Digital Land Cover Model for Germany (German Digitales Landbedeckungsmodell für Deutschland, LBM-DE, until 2012 DLM-DE) contained areal land cover information based on the CORINE Land Cover European nomenclature (Hovenbitzer et al. 2014). The LBM-DE dataset comprises selected areal object types of the ATKIS® Basis-DLM (German Digitales Landbedeckungsmodell, digital land cover model) land cover model for settlements, transport, vegetation, and water bodies. The classes were adapted in a modified form to meet the specific requirements of the CLC. The minimum mapping area of the dataset was 1 ha. The dataset has been updated every three years to the respective reference year since the area-wide initial coverage in 2009, using multitemporal satellite image data (mainly RapidEye (post-2020), Sentinel-2, and orthophotos). Since 2012, land cover and land use have been recorded separately, with subsequent automatic transformation into the CLC nomenclature. While land cover was recorded using image data, the ATKIS® base DLM land use model of the respective reference year was the land use source of information.

3.2.2 Harmonization of the settlement data

Table 1 presents an overview of the land cover data and associated characteristics, which were utilized as inputs for the SLEUTH UGMr. Figure 4 illustrates the processing scheme. First, the OSM and LBM-De vector data were rasterized (30 m resolution). All datasets were then projected to ETRS89-extended / LAEA Europe and clipped to the extent of the metropolitan area of Ruhr. Every dataset was reclassified to generate boolean land cover input data, with urban areas and non-urban areas. The respective settlement classes and years associated with the specific data sources are shown in Table 1.

3.2.3 Slope, restriction areas, and transport

In addition to an urban dataset, SLEUTH UGMr requires spatially explicit information on slopes, restriction areas, and transport information. Slope information was derived using the 30 m digital elevation model of the Shuttle Radar Topography Mission (Farr et al. 2007). A dataset

Tab. 1: Settlement datasets and the utilized settlement classes

Data set	Settlement features and years	Projection/ Resolution
OSM via Ohsome API (Source: Raifer et al. 2019)	residential, industrial, commercial, retail, harbor, port, railway, lock, marina, quarry, construction, landfill, brownfield, stadium, recreation ground, golf course, sports center, common, allotments, playground, pitch, village green, cemetery, park, zoo, track, garden; 2012–2022.	WGS84, vector
WSF-EVO (Source: Marconcini et al. 2021)	0 = non-settlement in the WSF2015 Multitemporal retrospective classification of Landsat-5/7; 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1999, 1998, 1997, 1996, 1995, 1994, 1993, 1992, 1991, 1990, 1989, 1988, 1987, 1986, and 1985.	WGS84, 30 m
GHS_BUILT_LDSMT _GLOBE_R2018A _3857_30_V2_0 (Source: Corbane et al. 2019 & 2018, EC JCR et al. 2019)	Multitemporal classification of built-up presence. 0 = no data 1 = water surface 2 = land no built-up in any epoch 3 = built-up from 2000 to 2014 epochs 4 = built-up from 1990 to 2000 epochs 5 = built-up from 1975 to 1990 epochs 6 = built-up up to 1975 epoch	Mollweide, 30 m
LBM-DE (Source: ©GeoBasis-DE / BKG 2022)	1.1.1. Continuous urban fabric 1.1.2. Discontinuous urban fabric 1.2.1. Industrial or commercial units 1.2.2. Road and rail networks and associated land 1.2.3. Port areas 1.2.4. Airports 1.3.1. Mineral extraction sites 1.3.2. Dump sites 1.3.3. Construction sites 1.4.1. Green urban areas 1.4.2. Sport and leisure facilities. 2009, 2012, 2015, and 2018	ETRS89-extended / LAEA Europe, vector

on restricted areas was created by adding water body information from the LBM-DE dataset to the authoritative map on natural reserves from La ndschaftsinformationssammlung (LINFOS) NRW (2022). Transport information was derived using Geofabrik's download server of OSM data and the key "highway" (GEOFABRIK 2022). The three datasets were processed to meet UGMr standards (Rienow & Goetzke 2015). Table 2 presents the information required for the initialization of the modeling process. The streets were weighted with different values, representing the variation in different speeds on a particular road types (Goetzke 2012).

3.3 Areas affected by floods and thermal conditions

The analyses of the allocation of new urban areas within areas affected by floods and heat stress were based on two administrative datasets of the State of North Rhine-Westphalia (NRW) in Germany. The NRW flood hazard maps provided information on which areas are subjected to flooding and the associated water depths and flow velocities. One product is the "HQfrequent" (Hochwasser, German for floods, and Q is the discharge index) map representing relatively frequent floods that occur every 10 to 20 years, on

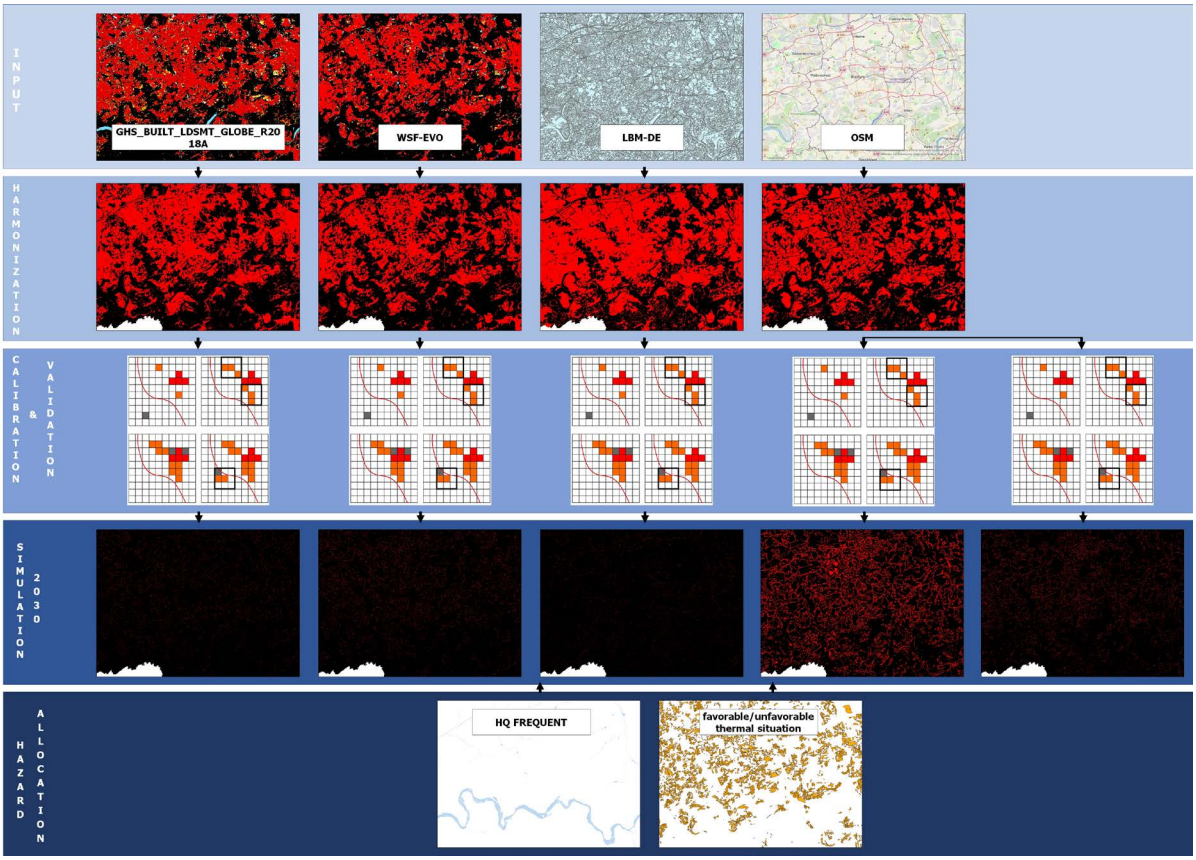

Fig. 4: Processing scheme and workflow of the study

average (for more information, refer to MUNLV 2022). The NRW state-wide climate analysis was carried out in 2018 to analyze the thermal and air-hygienic situation, as well as the effects of building and planning measures. Primarily, the spatial characteristics of air exchange and thermally stressed areas were considered, and the relationship between compensation and affected areas was investigated. Recommendations for planning to improve or maintain the situation were derived from these results. The overall assessment aimed to provide an integrated evaluation of the facts presented in the day and night situations with regard to planning-relevant concerns. It classified the thermal conditions as from "very favorable" to "extreme unfavorable" (LANUV 2018).

Tab. 2: Spatial information needed to set up SLEUTH UGMr model

Name	Description
Slope	0–97 percent
Natural reserve areas and waterbodies (exclusion of urban development)	0: urbanization possible 1: urbanization not possible
Transport	0: bridleway, cycleway, footway, path, pedestrian, steps, track, track grade 1-5; 25: living street, service 50: residential, tertiary, tertiary link; 100: motorway, motorway link, primary, primary link, secondary, secondary, link, trunk, trunk link

4 Results

4.1 Mapped urban areas and urban development in the Ruhr

Figure 5 depicts how heterogenous the data sets are when it comes to settlement mapping. The map depicts all areas labelled as "urban" in the different data sets. It shows those areas which are mapped as "urban" in all data sets, in some of the data sets, and if there are areas mapped as urban in just one of the data sets. Additionally, it comprises the quantitative settlement development of Ruhr between 1975 and 2022, based on one national and three global datasets. For example, having a look in the Eastern part at the city of Dortmund (Fig. 5 bottom) for 2014/15, we see that the OSM data set misses the whole downtown area. The urban coverage increases by 2022 but is still lacking (Fig. 5 c). Regarding the Ruhr this is the most extreme example of urban misses in the OSM data set. The other downtown districts are well covered. One can also observe the potential misses and false-positives of the other three data sets. The related land uses comprise cemeteries, horse race tracks, brown lands, and railroad areas. The settlement growth depicted as 95 km^2 (OSM data, for 2021–2022), 164 km^2 (WSF Evolution for 1985–2015), 249 km^2 (GHSL data, for 1975–2014), and 24 km^2 (LBM-DE, for 2015–2012). Different settlement extensions were observed, reaching 1,104 km^2 in the WSF Evolution data to 1,574 km^2 for the LBM-DE, whereas OSM and GHSL showed similar maximum values of 1,227 km^2 and 1,234 km^2, respectively, but in different years. Nevertheless, because the SLEUTH UGMr is trained with each dataset, a certain degree of consistency is guaranteed owing to the calibration and independent validation which are carried out for data from the same data source.

4.2 SLEUTH UGMr calibration

The temporal resolution of the global urban datasets utilized was heterogeneous. To sustain a certain stability and consistency, it was ensured that a comparable time span could be used for the calibration of the SLEUTH UGMr, while having a third independent dataset for the CA validation (Fig. 6). To achieve this, a minimum calibration period was chosen as 2012–2014, as it was covered by all datasets. Table 3 shows the growth coefficients and MRV mean factors of agreement for all resolutions (F_l) of the calibrated SLEUTH UGMr models. It presents distinct calibration periods and the results of the calibration process. The three most relevant coefficients for driving the SLEUTH UGMr (Clarke et al. 1997b) are also depicted. Since the scale is 0–100 the calibrated values are quite low compared to other regions of the world which experience higher urbanization rates (Chaudhuri & Clarke 2013). The spread coefficient (representing the growth of urban edges), controls the distribution of the largest portion of newly urbanized pixels. The GHSL depicts the lowest amount of edge growth (spread value 4), whereas the OSM-based models depict the highest number of newly urbanized cells in the neighborhood of existing cells (spread value 19). The OSM-based model calibrated from 2012 to 2017 must be used with caution, as the mapping activities have only recently begun (Fig. 3). This means that new urban areas should be simulated, but although they can already be observed in reality, they are not yet in the database. The more recent the OSM data, the more reliable it becomes. However, all the models showed MRV values > 95%.

4.3 SLEUTH UGMr validation

Table 4 presents the validation results for the SLEUTH model. The datasets used for validation were not used in the calibration process. Accordingly, the validation periods differed from the calibration periods, but the calibrated coefficients were maintained. For the GHSL-based CA, 10 additional independent years could be used through an earlier start year than calibrated and 14 years overlap with the calibration period (Fig. 6). For the WSF-based CA, five additional independent years could be used through an earlier start year than calibrated, but there was no overlap with the calibration period. For the LBM-DE-based CA, three additional independent years could be used through a later end year than calibrated, and three years overlapped with the calibration period. For OSM I-based CA, five additional independent years could be used through a later start year than calibrated, and there was no overlap with the calibration period. Finally, for the OSM II-based CA, five additional independent years could be used through an earlier start year than calibrated, but there was no overlap with the calibration period (Fig. 6). Barring the F_t-value utilized, the MRV was used to compare the simulated map with a "null model," which is a map containing the initial settlement pattern of each validation start year. Hence, it was considered to be map of mere urban persistence

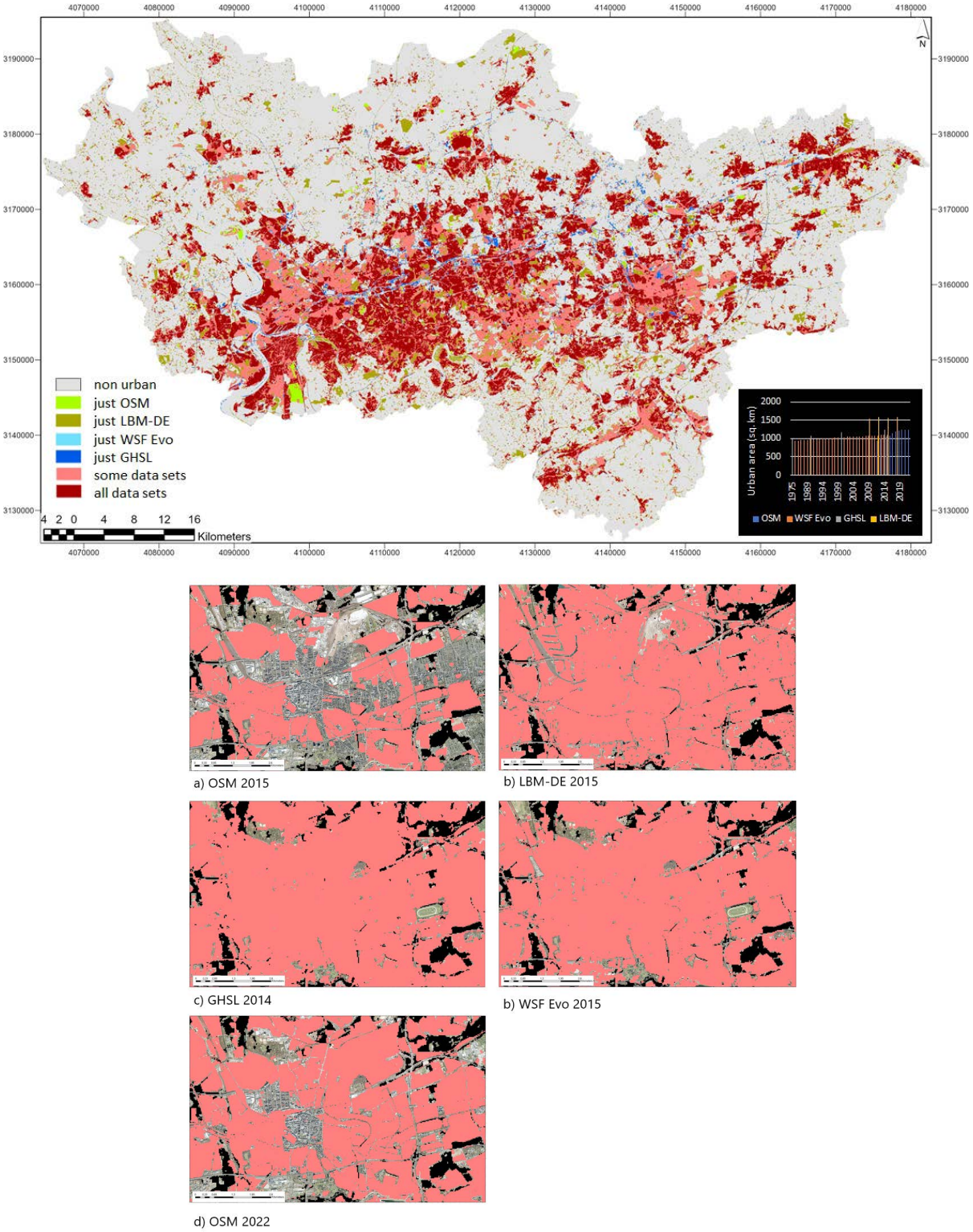

Fig. 5: Top: Comparison of urban classes in the different data sets (base year 2014/2015) and settlement development of the Ruhr 1975–2022 accordingly. Bottom: Urban coverage and potential misses (transparent) by the particular data sets 2014/2015 in the city of Dortmund. The OSM data set 2022 (d) reflect the ongoing mapping activities (base map: digital orthophoto 2015, Geobasis NRW).

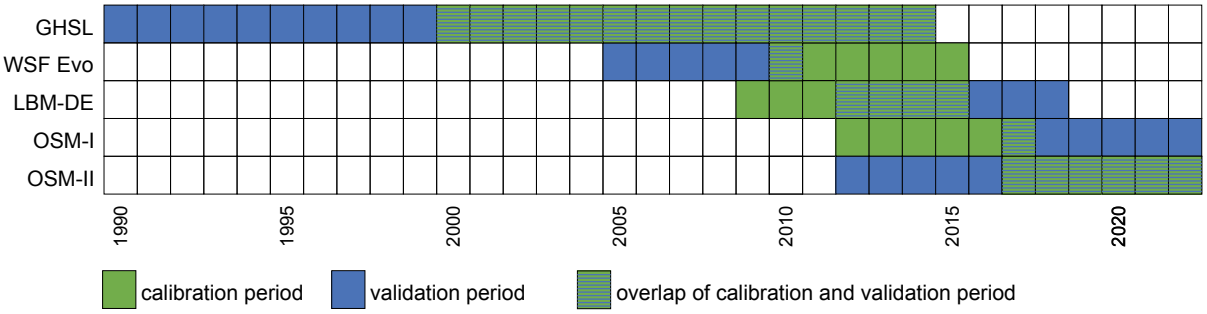

Fig. 6: Calibration and validation periods

Tab. 3: Growth coefficients and MRV mean factor of agreement over all resolutions (F_t)

Coefficient	GHSL (2000–2014)	WSF Evo (2010–2015)	LBM-DE (2009–2015)	OSM I (2012–2017)	OSM II (2017–2022)
Spread	4	7	12	63	19
Dispersion	7	10	4	66	11
Breed	27	13	11	69	20
$F_{t\ calibration}$	0.989	0.995	0.975	0.962	0.969

Tab. 4: Validation indices of the SLEUTH UGMr models

Validation index	GHSL (1990–2014)	WSF Evo (2005–2010)	LBM-DE (2012–2018)	OSM I (2017–2022)	OSM II (2012–2017)
$F_{t\ validation}$	0.961	**0.990**	0.980	**0.916**	0.955
MRV null resolution (m)	1,920	3,840	**122,880**	**30**	15,360
Kappa	0.887	**0.971**	0.954	**0.793**	0.883
Fuzzy Kappa	0.937	**0.987**	0.968	**0.872**	0.917
Kappa quantity	0.968	**0.995**	0.987	**0.897**	0.991
Kappa location	0.917	**0.976**	0.967	**0.883**	0.891
TP	1,222,987	**1,175,306**	1,712,415	**1,264,185**	1,248,206
TN	3,485,923	**3,702,856**	**3,117,264**	3,238,753	3,453,980
FP	78,098	**30,934**	66,782	**321,530**	106,303
FN	**140,799**	**21,417**	36,643	108,636	124,615
sensitivity	**0.897**	**0.982**	0.979	0.921	0.909
false negative rate	**0.103**	**0.018**	0.021	0.079	0.091
specificity	0.978	**0.992**	0.979	**0.910**	0.970
false positive rate	0.022	**0.008**	0.021	**0.090**	0.030

with no urban growth in that period (Pontius et al. 2008, Pontius & Malizia 2004, Visser 2004). The resolution level where the agreement factor of the land-use model outperforms the null model for the first time, is called "null resolution." The higher the null resolution, the better the land-use model performance. All F_t-values were lower than those in the calibration runs, but still reached a level > 90%. The MRV results showed an overall accuracy with an excellent level of agreement. This can be attributed to the ability of all the models to predict the location and quantity of non-urban cell states extremely well. This is emphasized by the specificity values and false positive rates.

Overall, the WSF-based CA data demonstrated the best performance. The GHSL-based CA had the lowest sensitivity value and the highest false negative rate, but also covered the longest modeling period. The LBM-DE-based CA data barely outperformed its null model. The interpretation of the null model comparison was heavily dependent on the observed urban growth. The higher the observed growth, the more likely that the null model was outperformed, because the effect of locational errors was lower than the effect of quantity errors. In scenarios with low urban growth, the opposite occurred. The LBM-DE datasets showed a decrease in settlement areas for 2012–2015 and 2012–2017. Accordingly, the SLEUTH UGMr simulates only growth model, and it cannot simulate shrinkage. However, it reaches a good modeling accuracy when using the LBM-DE dataset. The OSM-I-based CA was the only model which reached a null resolution of one. Since the OSM mapping activities were still in their infancy in 2012, it might be that new settlements in 2015 were not new but rather only recently mapped. In addition to accuracy measurements, the uncertainty of CA needs to be assessed, such as for the UGMr (Clarke 2004, Timmermans 2003, Wegener 2011). A proven procedure is the application of 100 Monte Carlo simulation runs reflecting the probability of each cell to be urbanized during the geosimulation (Aerts et al. 2003, Wegener 2011). The models simulated the settlement patterns of 2030 100 times. Figure 7 shows a violin plot of the resulting distributions of selected urban cells. It ranges from 1 to 99, since 0 represents "not selected as urban after 100 MC simulation runs" and 100 represents "cell already urban in the start year of the simulation". The plot includes the minimum, maximum, sample median, first, and third quartiles, as well as the probability densities of the different models. The OSM-I model distributed more pixels in the same location than any other model after 100 runs. Simultaneously, it showed more randomly selected pixels (i.e., low Monte Carlo simulation run values) than the other models. In addition, the OSM-I model showed the lowest median, whereas the OSM-II model showed the highest. The only other model that performs as well as OSM-II was the WSF-based CA.

4.3 Settlement simulation 2030

Figure 8 depicts the allocation of new settlement areas in regions of HQ-frequent (flooding) areas, and in areas with potential for heat stress. It clearly varies between datasets. This indicates that geosimulation model outcomes utilized in spatial planning should avoid rigorous normative messages without discussing the input data that has been used in the models.

Table 5 shows the number of new settlements simulated by the CA SLEUTH UGMr, after calibration and validation, using different urban data input. The Ruhr covers an are of 4,439 km^2. The area of new settlements in future simulations reached between 40.77 km^2 to 477.91 km^2, demonstrating the effect of the coefficients used in calibration of different datasets. The number of new settlements differs in absolute numbers in all CAs and within areas affected by floods and heat stress. The percentage of new settlements simulated in areas affected by frequent floods was very similar in the models, including the OSM I-based model, which overestimated urban growth owing to the higher observed growth numbers from 2012 to 2015. However, this did not hold true for areas with a potential for heat stress.

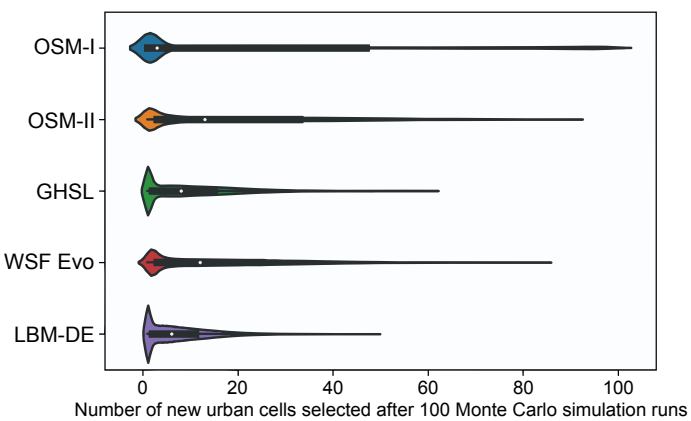

Fig. 7: Violin plot depicting the outcome 100 Monte Carlo simulation runs of the five CA

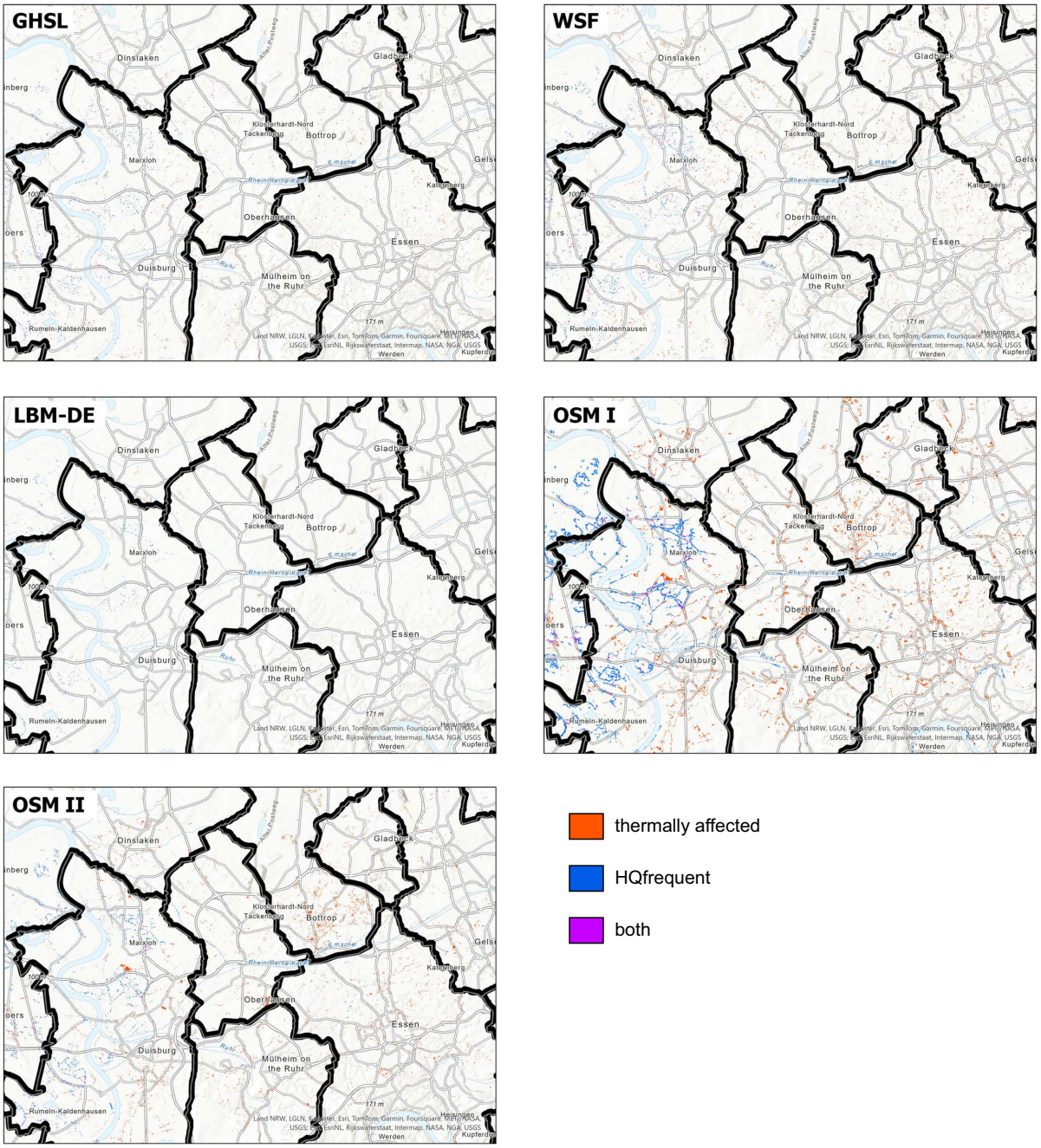

Fig. 8: The allocation of new settlement areas in regions of HQfrequent areas (i.e., flood affected areas), in areas with a potential for heat stress, and both HQfrequent and potential for heat stress depicted for the city of Duisburg in the Western Ruhr Area.

The LBM-DE-based SLEUTH UGMr showed the lowest absolute and relative values, while the OSM-I-based SLEUTH UGMr had the highest absolute value. The WSF Evo-based SLEUTH UGMr had the highest relative values, despite showing the lowest quantity of urban cells in the start and end years of the simulation.

5 Discussion

This study validated the allocation and quantity estimation capacities of OSM-based models compared to models based on global urban data sets and on a national data set (Fig. 6). The applied model, SLEUTH-UGM, is a cellular automaton-based urban

Tab. 5: Settlement growth simulated by the SLEUTH UGMr models and allocation 2030

	GHSL	WSF Evo	LBM-DE	OSM I	OSM II
Starting year of simulation	**2014**	2015	2018	**2022**	**2022**
Settlements observed (km²)	1,227.41	**1,103.96**	**1,574.15**	1,233.56	1,233.56
Settlements simulated (km²) 2030	1,284.73	**1,191.34**	1,614.92	**1,711.47**	1,354.75
Settlement growth simulated (km²) 2030	57.32	87.37	**40.77**	**477.91**	121.90
Settlement growth simulated (%) 2030	4.66	7.91	**2.25**	**38.74**	9.88
Settlement growth simulated in HQfrequent areas (km²) 2030	**3.10**	4.75	3.24	**32.08**	7.81
Settlement growth simulated in HQfrequent areas (%) 2030	**5.41**	5.43	**7.95**	6.71	6.44
Settlement growth simulated in areas with a potential for heat stress (sq. km) 2030	9.60	15.23	**0.25**	**40.49**	18.41
Settlement growth simulated in areas with a potential for heat stress (%) 2030	16.74	**17.42**	**0.62**	8.47	15.19
Settlement growth simulated in HQfrequent areas and areas with a potential for heat stress (sq. km) 2030	0.51	0.79	**0.03**	**1.98**	0.86
Settlement growth simulated in HQfrequent areas and areas with a potential for heat stress (%) 2030	0.89	**0.90**	**0.064**	0.415	0.70

Noet: Red values represent the minimum, blue values the maximum.

growth model widely used for predicting urban expansion. One of its key advantages lies in its simplicity and ease of implementation, particularly for regions with limited data availability or computational resources. SLEUTH relies on straightforward input parameters such as boolean urban information, slope, roads, and restricted areas are making it accessible to a broad range of users (Clarke et al. 2007). However, compared to other machine learning or deep learning-based approaches, SLEUTH has several limitations. Firstly, it may struggle to capture complex spatial interactions and non-linear relationships inherent in urban growth processes. Machine learning algorithms, such as random forests or neural networks, can often better accommodate these complexities, leading to more accurate predictions (Rienow et al. 2021). Secondly, SLEUTH's performance may degrade in rapidly changing urban environments or regions with unique characteristics not captured by its simplistic input parameters. In contrast, machine learning and deep learning models can adapt more flexibly to diverse contexts and evolving patterns, potentially yielding more robust predictions over time (Liu et al. 2020a).

The calibrated growth coefficients differed, leading to different growth rates. The datasets used for the calibration were from the same source; therefore, consistency can be guaranteed so that every model delivered the best performance based on the input urban datasets. Furthermore, overlap between the calibration and validation periods was avoided or reduced to the possible minimum. The data sets differed in terms of temporal resolution, coverage, and data end dates; therefore, an overlap between the datasets was required (2012–2014). Historic OSM data showed false-negative observations of urban growth owing to missing mapping activities, which lead to a potential overprediction of new urban areas in the future (Fig. 6). Accordingly, the study tested two calibration periods, which demonstrated the importance of choosing the correct start year for the OSM-based geosimulation. An important consideration in studies is the definition of what is considered to be urban and what is considered non-urban. The global earth observation-based datasets (GHSL and WSF Evolution) used in this study focused on urban land cover (i.e., primarily impervious surfaces). The administrative

data (LBMDE) and OSM-derived data also considered urban land uses, such as urban recreational areas and leisure facilities. Open spaces within an urban area were the first to experience edge expansion, which is the most dominant growth type of CA. This could explain why the WSF-and GHSL-based CA allocated more new urban pixels in areas with a less favorable / unthermal conditions, in 2030 simulations.

While SLEUTH-UGM is widely recognized for its effectiveness in simulating urban growth patterns, it's essential to acknowledge that no single model is universally superior in all contexts. CA models, for instance, are renowned for their simplicity and computational efficiency, making them suitable for large-scale applications. However, they often struggle to capture the complexities of urban processes, such as socioeconomic dynamics and land-use interactions. On the other hand, agent-based models (ABM) excel in representing individual behaviors and interactions, offering a more detailed and realistic portrayal of urban growth dynamics. Nonetheless, they can be computationally intensive and require extensive parameterization and validation efforts. However, it's crucial to recognize that each modeling approach has its strengths and limitations, and the choice of model should align with the specific research objectives and data constraints of the study area. Furthermore, integrating multiple modeling techniques, such as combining CA models with ABM, can offer synergistic benefits, enhancing the accuracy and comprehensiveness of urban growth predictions. Thus, while the findings provide valuable insights into the performance of SLEUTH-UGM, a holistic understanding of urban growth dynamics requires consideration of a diverse range of modeling approaches and methodologies.

6 Conclusion

This study calibrated, validated, and implemented an urban CA based different global LULC products and compared satellite-based, VGI-based, and administrative data sets for regional modeling purposes. The OSM-based CA and its simulation capacities were compared with two well-known satellite-based datasets, WSF Evo and GHSL, and a national administrative land cover dataset. The heterogeneous datasets have been harmonized so that the SLEUTH UGMr could be set up and the future urban growth of Ruhr was simulated until 2030. All models achieved high calibration and validation accuracy. The OSM-based CA was calibrated with the most recent datasets. The WSF-based CA showed a higher certainty than the other three Cas which were tested. The older the OSM data, the higher the probability of type-I errors (false positives). The global WSF-Evo dataset outperformed the national LBM-DE dataset. The latter met the null resolution only at a level where one pixel was approximately the same size as the entire study area. The four CAs behaved differently in terms of growth coefficients, leading to a variation in the quantity of future settlement growth between 40.77 km^2 (LBM-DE) and 477.91 km^2 (OSM-I). The higher the simulated urban growth, the higher the absolute number of new settlements within areas affected by environmental hazards, such as floods and heat stress. All models relatively allocated a similar size of new settlement areas affected by floods; however, they varied with respect to thermal conditions.

The study demonstrates the sensitivity of urban models when it comes to their input data, and in this study specifically, boolean global urban datasets. Environmental planning and hazard monitoring in metropolitan areas rely on predictions based on urban growth models. The key finding of this study is that scientifically evaluated and approved data sets still lead to different outcomes in urban modeling and may affect the measurements by evidence-based decision making. Future research needs to focus on the definition of "urban" in global products. For example, while some products exclusively incorporate impervious surfaces, others also incorporate inner urban green areas. In addition, it is necessary to know how the nature of urban input data and the associated potential future patterns might affect environmental evaluation studies and results.

Data availability

The data that support the findings of this study are openly available in RUB GitLab repository "Forecasting Urban Futures – Data" at https://git-lab.ruhr-uni-bochum.de/rienoat4/forecasting-urban-futures-data, reference number 3170.

Acknowledgment

The authors thank the German Federal Agency for Cartography and Geodesy for providing the LBM-DE datasets.

References

Adamo SA, Baker JL, Lovett MME, Wilson G (2012) Climate change and temperate zone insects: The tyranny of thermodynamics meets the world of limited resources. *Environmental Entomology* 41: 1644–1652. https://doi.org/10.1603/EN11188

Aerts CJH, Clarke KC, Keuper AD (2003) Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science* 30: 249–261. https://doi.org/10.1559/152304003100011180

Badmos OS, Rienow A, Callo-Concha D, Greve K, Jürgens C (2019) Simulating slum growth in Lagos: An integration of rule based and empirical based model. *Computers, Environment and Urban Systems* 77: 101369. https://doi.org/10.1016/j.compenvurbsys.2019.101369

Batty M (2008) Fifty years of urban modeling: Macro-statics to micro-dynamics. Albeverio S, Andrey D, Giordano P, Vancheri A (eds): *The dynamics of complex urban systems:* 1–20. Heidelberg, New York. https://doi.org/10.1007/978-3-7908-1937-3_1

Benenson I, Torrens PM (2004) Geosimulation: Automata-based modeling of urban phenomena. Chichester. https://doi.org/10.1002/0470020997

Chaudhuri G, Clarke KC (2013) The SLEUTH land use change model: A review. *International Journal of Environmental Resources Research* 1: 88–104.

Chien YMC, Carver S, Comber A (2020) Using geographically weighted models to explore how crowdsourced landscape perceptions relate to landscape physical characteristics. *Landscape and Urban Planning* 203: 103904. https://doi.org/10.1016/j.landurbplan.2020.103904

Clarke KC (2004) The limits of simplicity: Toward geocomputational honesty in urban modeling. *Peter Atkinson P, Foody GM, Darby SE, Wu F (eds) GeoDynamics:* 215–232. Boca Raton. https://doi.org/10.1201/9781420038101.ch16

Clarke KC (2008) Mapping and modelling land use change: An application of the SLEUTH model. Pettitt C, Cartwright W, Bishop I, Lowell K, Pullar D, Duncan D (eds): *Landscape analysis and visualization. Lecture notes in geoinformation and cartography:* 353–366. Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69168-6_17

Clarke KC, Gazulis N, Dietzel CK, Goldstein N (2007) A decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model. Fischer P (ed) *Classics from IJGIS. Twenty years of the International Journal of Geographical Information Systems and Science*: 413–425. London.

Clarke KC, Hoppen S, Gaydos L (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design* 24: 247–262. https://doi.org/10.1068/b240247

Clarke KC, Johnson JM (2020) Calibrating SLEUTH with big data: Projecting California's land use to 2100. *Computers, Environment and Urban Systems* 83: 101525. https://doi.org/10.1016/j.compenvurbsys.2020.101525

Corbane C, Florczyk A, Pesaresi M, Politis P, Syrris V (2018a) GHS-BUILT R2018A - GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014) - OBSOLETE RELEASE. https://doi.org/10.2905/jrc-ghsl-10007

Corbane C, Florczyk A, Pesaresi M, Politis P, Syrris V (2018b) GHS-BUILT R2018A - GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014) - OBSOLETE RELEASE. https://doi.org/10.2905/jrc-ghsl-10007

Corbane C, Pesaresi M, Kemper T, Politis P, Florczyk AJ, Syrris V, Melchiorri M, Sabo F, Soille P (2019) Automated global delineation of human settlements from 40 years of Landsat satellite data archives. *Big Earth Data* 3: 140–169. https://doi.org/10.1080/20964471.2019.1625528

Costanza R, Maxwell T (1991) Spatial ecosystem modelling using parallel processors. *Ecological Modelling* 58: 159–183. https://doi.org/10.1016/0304-3800(91)90034-X

EC JRC (European Commission, Joint Research Centre), Freire S, Corbane C, Zanchetta L, Schiavina M, Politis P, Kemper T, Ehrlich D, Pesaresi M, Maffenini L, Florczyk AJ, Melchiorri M, Sabo F (2019) GHSL data package 2019: Public release GHS P2019. Luxembourg. https://human-settlement.emergency.copernicus.eu/documents/GHSL_Data_Package_2019.pdf

EC JRC (European Commission, Joint Research Centre) (2023) GHSL data package 2023. Luxembourg. https://doi.org/10.2760/098587

Else H (2021) Climate change implicated in Germany's deadly floods. *Nature*. https://doi.org/10.1038/d41586-021-02330-y

Estima J, Painho M (2015) Investigating the potential of OpenStreetMap for land use/land cover production: A case study for continental Portugal. Jokar Arsanjani J, Zipf A, Mooney P, Helbich M (eds): *OpenStreetMap in GIScience: Experiences, research, and applications. Lecture Notes in Geoinformation and Cartography:* 273–293. Cham. https://doi.org/10.1007/978-3-319-14280-7_14

European Commission (2023) GHSL - Global Human Settlement Layer. https://ghsl.jrc.ec.europa.eu/download.php?ds=sdata

Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf D (2007) The Shuttle Radar Topography Mission. *Reviews of Geophysics* 45. https://doi.org/10.1029/2005RG000183

Florczyk A, Politis P, Corbane C, Pesaresi M (2018) GHS-BUILT R2018A - GHS built-up grid INPUT DATA, Landsat multitemporal collections (1975-1990-2000-2014) - OBSOLETE RELEASE.

Fonte CC, Patriarca JA, Minghini M, Antoniou V, See L, Brovelli MA (2019) Using OpenStreetMap to create land use and land cover maps: Development of an application. Campelo CEC, Bertolotto M, Corcoran P (eds) *Geospatial intelligence: Concepts, methodologies, tools, and applications:* 1100–1123. https://doi.org/10.4018/978-1-5225-8054-6.ch047

GEOFABRIK (2022) https://www.geofabrik.de/de/index.html.

Goetzke R (2012) Entwicklung eines fernerkundungsgestützten Modellverbundes zur Simulation des urban-ruralen Landnutzungswandels in Nordrhein-Westfalen. Hamburg.

Hassan MI, Elhassan SMM (2020) Modelling of urban growth and planning: A critical review. *Journal of Building Construction and Planning Research* 8: 245–262. https://doi.org/10.4236/jbcpr.2020.84016

Hospers G-J, Wetterau B (2018) Small Atlas Metropole Ruhr: The Ruhr Region in transformation. Essen. https://doi.org/10.13140/RG.2.2.16724.04480

Hovenbitzer M, Emig F, Wende C, Arnold S, Bock M, Feigenspan S (2014) Digital land cover model for Germany – DLM-DE. Manakos I, Braun M (eds): *Land use and land cover mapping in Europe: Practices & trends, remote sensing and digital image processing:* 255–272. Dordrecht. https://doi.org/10.1007/978-94-007-7969-3_16

Jantz CA, Goetz SJ, Donato D, Claggett P (2010) Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Computers, Environment and Urban Systems* 34: 1–16. https://doi.org/10.1016/j.compenvurbsys.2009.08.003

Jokar Arsanjani J, Zipf A, Mooney P, Helbich M (2015) An introduction to OpenStreetMap in geographic information science: Experiences, research, and applications. Jokar Arsanjani J, Zipf A, Mooney P, Helbich M (eds) *OpenStreetMap in GIScience: Experiences, research, and applications. Lecture notes in geoinformation and cartography:* 1–15. Cham. https://doi.org/10.1007/978-3-319-14280-7_1

Judex M (2008) Modellierung der Landnutzungsdynamik in Zentralbenin mit dem XULU-Framework. Bonn.

Krelaus L, Apfel J, Rienow A (2021) Satellite-based investigation on the surface cooling effects of urban parks and their range – a case study for North Rhine-Westphalia, Germany. *Erdkunde 75*: 209–223. https://doi.org/10.3112/erdkunde.2021.03.03

Landschaftsinformationssammlung (LINFOS) NRW (2022) https://www.opengeodata.nrw.de/produkte/umwelt_klima/naturschutz/linfos/

LANUV (Federal State Office for Nature, the Environment and Consumer Protection in North Rhine Westphalia) (2018) Klimaanalyse Nordrhein-Westfalen. Recklinghausen. https://www.lanuv.nrw.de/klima/klimaanpassung-in-nrw/klimaanalyse

Liu D, Clarke KC, Chen N (2020a) Integrating spatial nonstationarity into SLEUTH for urban growth modeling: A case study in the Wuhan metropolitan area. *Computers, Environment and Urban Systems* 84: 101545. https://doi.org/10.1016/j.compenvurbsys.2020.101545

Liu F, Wang S, Xu Y, Ying Q, Yang F, Yuchu Q (2020b) Accuracy assessment of Global Human Settlement Layer (GHSL) built-up products over China. *PLOS ONE* 15: e0233164. https://doi.org/10.1371/journal.pone.0233164

Marconcini M, Metz-Marconcini A, Esch T, Gorelick N (2021) Understanding current trends in global urbanisation - The World Settlement Footprint suite. *GI_Forum* 9: 33–38. https://doi.org/10.1553/giscience2021_01_s33

Marconcini M, Metz-Marconcini A, Üreyen S, Palacios-Lopez D, Hanke W, Bachofer F, Zeidler J, Esch T, Gorelick N, Kakarla A, Paganini M, Strano E (2020) Outlining where humans live, the World Settlement Footprint 2015. *Scientific Data* 7: 242. https://doi.org/10.1038/s41597-020-00580-5

MUNLV (Ministry of Agriculture and Consumer Protection of the State of North Rhine-Westphalia) (2022) Hochwassergefahrenkarten - Inhalte und Symbole. Flussgebiete NRW. https://www.flussgebiete.nrw.de/hochwassergefahrenkarten-inhalte-und-symbole

Oleson KW, Monaghan A, Wilhelmi O, Barlage M, Brunsell N, Feddema J, Hu L, Steinhoff DF (2015) Interactions between urbanization, heat stress, and climate change. *Climatic Change* 129: 525–541. https://doi.org/10.1007/s10584-013-0936-8

Patriarca J, Fonte CC, Estima J, de Almeida J-P, Cardoso A (2019) Automatic conversion of OSM data into LULC maps: Comparing FOSS4G based approaches towards an enhanced performance. *Open Geospatial Data, Software and Standards* 4: 11. https://doi.org/10.1186/s40965-019-0070-2

Pesaresi M, Politis P (2023) GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset]. https://doi.org/10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA

Pesaresi M, Syrris V, Julea A (2016) A new method for earth observation data analytics based on symbolic machine learning. *Remote Sensing* 8: 399. https://doi.org/10.3390/rs8050399

Pontius RG, Boersma W, Castella J-C, Clarke K, Nijs T, Dietzel C, Duan Z, Fotsing E, Goldstein N, Kok K, Koomen E, Lippitt C, McConnell W, Sood AM, Pijanowski B, Pithadia S, Sweeney S, Trung T, Veldkamp A, Verburg P (2008) Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science* 42: 11–37. https://doi.org/10.1007/s00168-007-0138-2

Pontius RG, Huffaker D, Denman K (2004) Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling* 179: 445–461. https://doi.org/10.1016/j.ecolmodel.2004.05.010

Pontius RG, Malizia NR (2004) Effect of category aggregation on map comparison. Egenhofer MJ, Freska C, Miller HJ (eds) *Geographic Information Science. Third International Conference, GI Science 2004 Adelphi, MD, USA, October 20-23, 2004 Proceedings*: 251–268. https://doi.org/10.1007/978-3-540-30231-5_17

Rafiee R, Mahiny AS, Khorasani N, Darvishsefat AA, Danekar A (2009) Simulating urban growth in Mashad City, Iran through the SLEUTH model (UGM). *Cities* 26: 19–26. https://doi.org/10.1016/j.cities.2008.11.005

Raifer M, Troilo R, Kowatsch F, Auer M, Loos L, Marx S, Przybill K, Fendrich S, Mocnik F-B, Zipf A (2019) OSHDB: A framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospatial Data, Software and Standards* 4: 3. https://doi.org/10.1186/s40965-019-0061-3

Rienow A, Goetzke R (2015) Supporting SLEUTH – Enhancing a cellular automaton with support vector machines for urban growth modeling. *Computers, Environment and Urban Systems* 49: 66–81. https://doi.org/10.1016/j.compenvurbsys.2014.05.001

Rienow A, Kantakumar LN, Ghazaryan G, Dröge-Rothaar A, Sticksel S, Trampnau B, Thonfeld F (2022) Modelling the spatial impact of regional planning and climate change prevention strategies on land consumption in the Rhine-Ruhr Metropolitan Area 2017–2030. *Landscape and Urban Planning* 217: 1–20. https://doi.org/10.1016/j.landurbplan.2021.104284

Rienow A, Mustafa A, Krelaus L, Lindner C (2021) Modeling urban regions: Comparing random forest and support vector machines for cellular automata. *Transactions in GIS* 25: 1625–1645. https://doi.org/10.1111/tgis.12756

Rienow A, Stenger D (2014) Geosimulation of urban growth and demographic decline in the Ruhr: A case study for 2025 using the artificial intelligence of cells and agents. *Journal of Geographical Systems* 16: 311–342. https://doi.org/10.1007/s10109-014-0196-9

Rienow A, Stenger D, Menz G (2014) Sprawling cities and shrinking regions - Forecasting urban growth in the RUHR for 2025 by coupling cells and agents. *Erdkunde* 68: 85–107. https://doi.org/10.3112/erdkunde.2014.02.02

Saxena A, Jat MK, Clarke KC (2021) Development of SLEUTH-Density for the simulation of built-up land density. *Computers, Environment and Urban Systems* 86: 101586. https://doi.org/10.1016/j.compenvurbsys.2020.101586

Schmitz M, Bode T, Thamm H-P, Cremers AB (2007) XULU - A generic JAVA-based platform to simulate land use and land cover change (LUCC). Oxley L, Kulasiri D (eds) *MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand*, December 2007: 2645–2651. https://www.mssanz.org.au/MODSIM07/papers/46_s60/XULU-AGenerics60_Schmitz_.pdf

Schultz M, Voss J, Auer M, Carter S, Zipf A (2017) Open land cover from OpenStreetMap and remote sensing. *International Journal of Applied Earth Observation and Geoinformation* 63: 206–213. https://doi.org/10.1016/j.jag.2017.07.014

Siedentop S, Fina S, Krehl A (2014) Greenbelts in Germany's regional plans – An effective growth management policy? *Landscape and Urban Planning* 145: 71–82. https://doi.org/10.1016/j.landurbplan.2015.09.002

Timmermans H (2003) The saga of integrated land use-transport modeling: How many more dreams before we wake up. Axhausen KW (ed) Moving through nets: The physical and social dimensions of travel. Selected papers from the 10th international conference on travel behaviour research. Luzern. https://doi.org/10.2208/jscej.2003.744_163

Visser H (2004) The Map Comparison Kit:. Methods, software, and applications. Maastricht.

Wang Z, Bachofer F, Koehler J, Huth J, Hoeser T, Marconcini M, Esch T, Kuenzer C (2022b) Spatial modelling and prediction with the spatio-temporal matrix: A study on predicting future settlement growth. *Land* 11: 1174. https://doi.org/10.3390/land11081174

Wang J, Rienow A, David M, Albert C (2022a) Green infrastructure connectivity analysis across spatiotemporal scales: A transferable approach in the Ruhr Metropolitan Area, Germany. *Science of the Total Environment* 813: 152463. https://doi.org/10.1016/j.scitotenv.2021.152463

Wang Z-H, Upreti R (2019) A scenario analysis of thermal environmental changes induced by urban growth in Colorado River Basin, USA. *Landscape and Urban Planning* 181: 125–138. https://doi.org/10.1016/j.landurbplan.2018.10.002

Wegener M (2011) From macro to micro. How much micro is too much? *Transport Reviews* 31 2: 161–177. https://doi.org/10.1080/01441647.2010.532883

Wu F (1998) SimLand: A prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of*

Geographical Information Science 12: 63–82. https://doi.org/10.1080/136588198242012

Zhou Y, Ma T, Zhou C, Xu T (2015) Nighttime light derived assessment of regional inequality of socioeconomic development in China. *Remote Sensing* 7: 1242–1262. https://doi.org/10.3390/rs70201242

Zhou Y, Varquez, ACG, Kanda, M (2019) High-resolution global urban growth projection based on multiple applications of the SLEUTH urban growth model. *Science Data* 6: 34. https://doi.org/10.1038/s41597-019-0048-z

Author

Jun. Prof. Dr. Andreas Rienow
ORCID: 0000-0003-3893-3298
andreas.rienow@rub.de
Ruhr University Bochum
Institute of Geography
Geomatics Research Group
Universitätsstraße 150
44801 Bochum
Germany