SLEUTHING AROUND MAP SIZES

-

EXPLORING THE SENSITIVITY TO SPATIAL EXTENT CHOICE IN MODELING URBANIZATION

MSc THESIS

Fidél Csikós

29th of June, 2021

Supervised by dr. Eduardo Simao Da Graca Dias

ABSTRACT

This thesis investigates the SLEUTH model's sensitivity to changes in spatial extent, i.e. the map size of the studied area. SLEUTH is a cellular automata-based urban growth model, which since its creation in 1997 has been used in more than a hundred studies to model urbanization worldwide. Despite its widespread usage, some of the model's core aspects are still not well understood - one such area is the effect of map size choice on the model outcome. This study attempts to promote discussion on this topic by modeling the urbanization around the Dutch city of Almere by four different map sizes and comparing the results. The different aspects of spatial sensitivity are examined by comparing the model outcomes, including the SLEUTH coefficients – a set of integers describing the underlying drivers of urbanization – and the spatial fit produced by the different map sizes. The results show that even slight differences in spatial extent can result in significant differences regarding the model outcome. Careful analysis can yield more insight into the divergent results, but this paper argues that for such inference it is necessary to take into consideration the context of the experiment and the intricate mechanisms of SLEUTH as well. This paper concludes that even though the effect of spatial extent is an often overlooked aspect of the modeling exercise, given its profound effect on model outcome, further research is needed on the topic.

TABLE OF CONTENTS

INTRODUCTION 2
Background and inspiration 2
Research question 2
Research design 3
Positioning in relevant literature 6
DISCUSSION 7
The SLEUTH model 7
Experiment setup 13
Choosing a city to model 13
Choosing a time period for study 15
Choosing a spatial extent 16
Measuring spatial fit 20
Data preparation 22
Calibration phase 27
Prediction phase 32
Results of the prediction runs 32
Question 1 – differences in the coefficients 34
Question 2 – differences of overall spatial fit 36
Question 3 - differences regarding the common extent 41
CONCLUSION 42
Summary of findings 42
Core observations 42
Stray observations 43
Limitations 45
Further studies 46
REFERENCES 48
APPENDIX 51

INTRODUCTION

Background and inspiration

This thesis is the continuation of the explorative analysis carried out as part of a Spatial Economics Research Project. It investigated the relationship between the level of urban planning and the coefficients of the SLEUTH model – a set of five integers driving the model's behavior –, concluding that even though there is strong evidence that a correlation between patterns of urban planning and SLEUTH coefficients exists, looking at the coefficients alone is not enough to compare urbanization trends across cities. The paper suggested that for credible comparisons, a wider range of model parameters – such as model version, spatial resolution, and size – should be accounted for as well.

One of the striking realizations after reviewing a wide array of studies utilizing SLEUTH to predict urban growth was the apparent arbitrariness in selecting the urban extent – map size – for modeling. The definition of a city extent is naturally subjective, requiring discretionary choices – for instance, whether to include suburban satellite settlements as part of the city under study. Nevertheless, due to the model's way of working it is reasonable to suspect that such decisions also influence the outcome of the model. But by what extent? Is there a way to find a way to select an optimal map size for the most accurate results? Even though hundreds of papers have been written either utilizing SLEUTH or evaluating its capabilities, no explicit investigation has been carried out on the effect of spatial extent on the model outcome. This paper attempts to kickstart the investigations in this direction.

Research question

This paper aims to focus attention on the effect of map size used for modeling, by investigating how sensitive SLEUTH is to spatial extent. This is driven by the overarching aspiration to provide guidance for selecting the 'optimal' spatial extent for a given scenario, acknowledging that optimality regarding spatial fit is a multifaceted issue, with different interpretations instead of a single quantifiable metric.

Therefore, three different aspects of sensitivity will be explored as sub-questions. First, it needs to be established whether and how does a change in spatial extent affects the model outcome after calibration – the coefficients describing the different drivers of urbanization. Second, whether larger or smaller map sizes tend to result in better overall fit. This can indicate the model's inherent capacity to pick up on urbanization trends at different extents but is alone not useful for future researchers focusing on a specific area, looking to choose an optimal map size around their area of interest. Lastly, for this purpose, the spatial fit of the intersect of different predictions made with different map sizes on the same area will be compared.

Research design

Due to the stochastic nature of SLEUTH – cellular automata-based simulations are carried out and averaged in a Monte Carlo-fashion – running simulations to gain insight into its capabilities is the first and foremost tool available for researchers (Clarke, 2008). This study utilizes this method as well – running simulations with the model on the same area with different map sizes and comparing the results – to investigate the stated questions. To achieve adequate results, challenges on three different subjects – computational, statistical, and geographical – needed to be tackled, which are all described in detail separately.

From the computational side, the main challenge was to optimize the model's performance so that satisfactory results can be achieved within the limited time and resources available for the study. The baseline version of the model – used in the overwhelming majority of studies – utilizes a brute-force method for calibration to derive the correct growth coefficients that drive the model's behavior. The main drawback of this method is the significant computational requirement – often measured in CPU days or weeks for larger SLEUTH projects. To combat this, this study initially utilized one of the latest advancements related to SLEUTH, the use of Genetic Algorithms (GA) to speed up the calibration process. GA provided significant gains in speed and comparable outcome in terms of spatial fit, but the results achieved suffered from very low robustness, making them unsuitable for the comparative analysis pursued by this paper. To guarantee reproducibility, a three-phase brute-force calibration method was used.

As this study compares predictions made with different spatial extents, the output is largely dependent on what statistical method is used for model selection and goodness of fit analysis. The optimal measurement for spatial fit is a highly discussed topic both within the SLEUTH and the larger land-use modeling community as well. SLEUTH itself provides a set of spatial fit statistics, but their application relies on the researcher. Multiple studies have been written on trying to find an optimal general metrics by either utilizing the default statistics or developing new measurement methods (Dietzel & Clarke, 2007; Jantz et al., 2010). For model selection and comparison of overall fit, this paper uses the Optimal Sleuth Metric (OSM) – a composite index of multiple spatial fit metrics, which has since proved to be a reliable tool for model selection. In the last part of the study, for the direct comparison of the intersects, the Lee-Salle metric is utilized due to its better interpretability. The model outputs urbanization results on a probability map, and this paper utilizes this advantage: for comparison, a 'probabilistic Lee-Salle metric' is calculated, which takes into consideration the chance of urbanization predicted for individual cells as well.

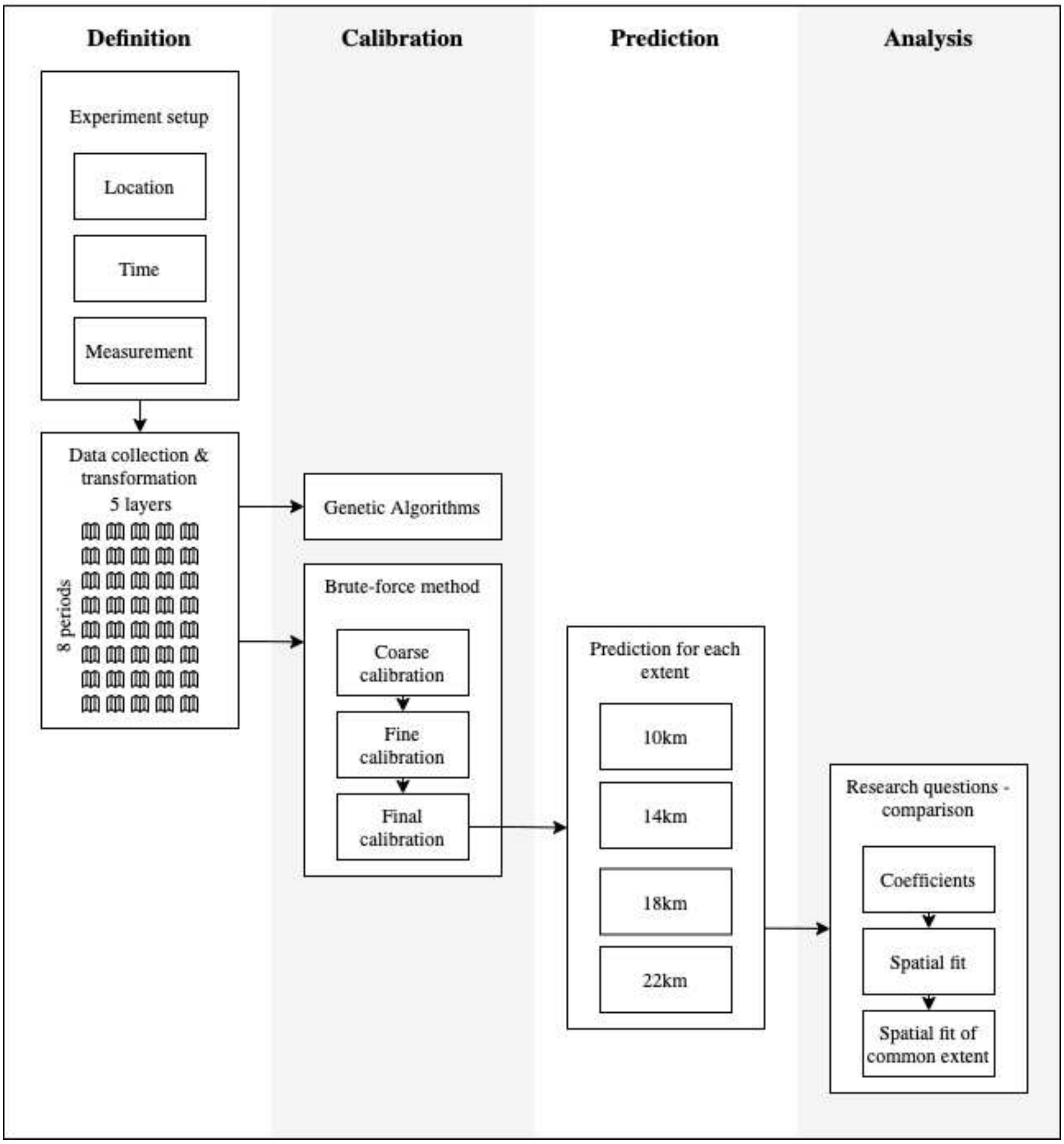
SLEUTH has very specific input requirements regarding the geographical data, therefore GIS-related tasks like data collection and preparation encompassed a significant part of the project. The Dutch city of Almere was selected for the modeling – as the newest city in the Netherlands built on reclaimed land, it is uniquely appropriate for modeling urban growth even within a shorter period. In line with recommendations by Dietzel and Clarke (2004), a relatively high spatial resolution of $25m^2$ is used to capture smaller-scale developments as well. The model requires rasterized maps for land slope, urban extent, transport network, etc. as input layers. These input layers have been prepared from three different data sources – CBS (Statistics Netherlands) Land Use maps, Copernicus Digital Elevation Maps (DEM), and OpenStreetMap transportation layers. The artifacts of different data sources have been reprojected and aligned to create coherent and valid inputs for the model.

The discussion of the paper follows a linear fashion. First, the model in general is introduced, followed by a detailed discussion of the setup of the experiments, along with the three clusters of challenges (computational, statistical, geographical) described. The next chapters detail the steps taken during the data preparation and calibration processes. It is worth highlighting that this study lays out the reasoning behind user-made choices and the steps taken to calibrate the model to a greater degree than is usual in SLEUTH literature. This is to assure reproducibility and comparison, inspired by the conclusions of the research project that was the forerunner of this thesis. The review of the literature using SLEUTH for modeling urban growth found that the details of the model setup are very rarely reported, leading to a weaker understanding of the results and almost no chance to reproduce the results.

Following these sections, the results of the experiments are discussed concerning the three aspects of sensitivity specified. First, the coefficient outcomes after calibration with different spatial extent are compared to see if the different extents resulted in meaningfully different coefficients. This is followed up by a comparison of extents based on their overall spatial fit. To investigate the source of possible differences, a unique 'topography of spatial fit' map is created to compare the spatial fit of different regions, such as core vs. periphery. Lastly, the common intersect of all the extents are compared by their spatial fit. The paper concludes with a summary, discussion of limitations, and notes on further studies.

Figure 1

Methodology of the modeling experiments



Positioning in relevant literature

The literature around SLEUTH can be grouped mostly into two categories: they are either written *using* it as a tool for application or they are written *about* SLEUTH, exploring its capacities. It is worth noting that there is a dynamic interaction between these two categories: new results using the model often inspire researchers to better understand and extend its capabilities – resulting in new developments which will then be incorporated into future models. Even though this paper does model land-use change in a specific region, it firmly belongs to the second group – it is written about SLEUTH, exploring its capabilities in the hope of enabling future researchers to make more informed decisions about the spatial extent being applied.

SLEUTH is a mature model with more than 20 years of history, which is a respectable age among land-use models. During this period it has seen many of its initial shortcomings improved upon or at least better understood, while some areas remain largely unexplored. Focusing on the spatial extent, this study aims to fill one of these holes. Initial research on SLEUTH mostly focused on (i) improving its computational efficiency, (ii) interpretation of the best spatial fit and (iii) the possibility to extend its capabilities by coupling it with other modeling tools or methods (Clarke, 2008). This paper greatly benefitted and uses the latest advances in the first two areas. After significant developments on these fronts, the last decade has seen a greater emphasis on trying to understand the relationship between the different parameters of the model and the coefficients it produces. This has been mainly driven by the aim to discover how to calibrate the model optimally, thus producing results with better spatial fit. Discussing still unsolved issues, the creator of the model, Clarke (2008) highlighted that the spatio-temporal sensitivity of the model remains largely unexplored.

Questions regarding temporal sensitivity have been first investigated in detail by Candau, who proved that SLEUTH produces more reliable results with a short history for a short forecast period than with a long history for a long forecast period. (2000; 2002). Chaudhuri and Clarke (2013) explored the temporal accuracy of the model further, but to this day, important practical questions like the optimal timing of a time step remain unanswered.

Studies around the spatial sensitivity of the model have largely focused on the effect of resolution on the model outcome. Noting the difficulty of finding the 'right' resolution, Dietzel and Clarke

(2004) note that finding the proper spatial resolution for the model is rather a tricky art and not yet a science – nevertheless, they propose a simple guideline for choosing a resolution based on the size of the area under study, which is followed by this paper as well. Further acknowledging the difficulty of addressing scale dependencies, Jantz and Goetz (2005) take an even more cautious approach, emphasizing the importance of considering the issue of resolution during modeling but without suggesting specific guidelines. These studies however only focused on one element of the spatial dimension, namely resolution. This paper investigates the other element: the effect of spatial size on model outcome.

DISCUSSION

The SLEUTH model

The forerunner of SLEUTH has been developed by Keith Clarke, a professor at the University of California during the 1990s, initially to model wildfire behavior. Even though he is still the most influential contributor for its development, at a very early phase his investigations have been transformed into a community effort: the source code was made available, and a website called Project Gigapolis[1] had been set up as a hub for the future development of the model. The initial version was published in 1997, while the latest iteration in 2017. Development has greatly benefitted from the concurrent experiments conducted by other researchers working on the model, whose achievements have been often included in the newer iterations of the model (Chaudhuri & Clarke, 2013). Such an example is the inclusion of Genetic Algorithms (GA) in the latest version of the model, which was inspired by other researchers' attempts to improve the model's efficiency by harnessing the capacities of GA – most notably by Noah Goldstein (2004).

SLEUTH is a cellular automata-based model - urban growth is simulated as a colony of unicellular organisms, where change is the outcome of the interaction between individual cells and the rules governing the growth of the colony. For SLEUTH, change happens in a set of nested loops: the outer loop represents the passage of time, while within the inner loop the rules governing the growth of the city (the 'colony' of cells) are executed in a sequential manner (Silva & Clarke, 2002). As the result of these rules, some new cells are born and some die at the end of each period: the new outcome will be used for calculating the growth in the next period, and so on. The simulation is always run for a specified number of periods, which usually correspond to years. SLEUTH

1 http://www.ncgia.ucsb.edu/projects/gig/

contains two tightly coupled models: the one originally developed called Urban Growth Model (UGM) and the Land Cover Deltatron Model (LDM). UGM models land-use change in a binary manner, meaning that a specific area is designated either as urban or non-urban. The LDM expands this dichotomy by making it possible to model land-use transitions with arbitrary categories too (Watkiss, 2008). Despite its extended capabilities, LDM has seen far fewer applications than UGM (Venema, 2016). This study utilizes UGM as well.

The model has different process flows – modes – among which the calibration and prediction modes are the most important. During the calibration phase, the model attempts to deduce the trends driving urbanization from the historical data it is provided, and distill them to five coefficients, which represent the intensity of each of these trends (Liu et al., 2019). This has traditionally been done by different brute force methods, where temporary maps are generated by the different variations of the coefficient space. These interim predictions are compared against the provided historical maps, and the coefficients that produced predictions with the best spatial fit against the actual maps will be used for forecasting. During the prediction mode, the coefficient set selected during the calibration phase is used to generate urban extent maps into the future. The exact method of calibration, measurement of spatial fit, and coefficient selection has been in the center of meta-research regarding SLEUTH since its inception (Clarke, 2007).

SLEUTH requires six different types of map layers as rasterized greyscale images as inputs – Slope, Land Use, Exclusion, Urban, Transport, and Hillshade – the model got its name after the initials of these layers (Hua et al., 2014). For consistent calibration, the model requires at least four Urban (for UGM) or Land Use (for LDM) layers for different periods. The Slope layer represents land elevation, Exclusion the areas that are not available for urban development, and Transport the transportation network that can be conducive to urban growth. As these factors are expected to change less often, one instance is sufficient for the whole simulation, although specific layers for each period might be used as well. The instances of these four layer categories – Slope, Exclusion, Transport, and Urban or Land Use – are used by the model as input to deduce and predict the urbanization trends of the simulated region. Lastly, the Hillshade layer is only used for representation, to provide a backdrop layer for the urbanization process as a visual cue.

Within the inner loop, four growth rules are sequentially executed. The first of which is the Spontaneous growth rule which accounts for random urbanization. This means that based on the

Diffusion[2] coefficient, there is a small chance that a non-urban pixel[3] will become urbanized. This growth effect, just as all others, is affected by the land slope of that area: the detrimental effect of non-flat land on urbanization is governed by the slope coefficient. Subsequently, the next phase models the chance of newly urbanized areas to become the catalysts of further growth, propagated by the urbanization of additional neighboring cells. Further growth of already existing or just newly formed spreading centers (defined as having three urban cells within the same 3×3 neighborhood) is accounted for in the following phase called the Edge growth phase. This phase represents the organic growth along the edges of the city. Lastly, the impelling effect of the proximity to the transportation network (usually roads) on urbanization is taken into account in the Road-influenced growth phase by spurring further growth along transportation networks (Clarke et al., 1997). The coefficients acting on the specific growth rules are listed in Table 1. To account for the often-observed boom and bust cycles of urban development, the model has a self-modifying algorithm to constantly update the initial coefficients within a pre-defined limit. The characteristics of the boom and bust behavior are governed by certain hyperparameters, which are set at the beginning of the calibration phase, and do not change during the fitting process (Saxena & Jat, 2019).

Table 1

The effect of coefficients on the different growth rules

Growth rule	Spontaneous	New Spreading centers	Edge	Road-influenced
Coefficients	Diffusion			Diffusion
		Breed		Breed
			Spread	
				Road-gravity
	Slope	Slope	Slope	Slope

Different regions have distinct profiles of urban growth – some are more sensitive to elevation, or concentrated along the existing transportation network, or have a higher level of urban planning, and so on. These attributes together define what Silva and Clarke (2004) referred to as the 'DNA of a region', which is distilled to five growth coefficients – all integers between 1 and 100 - in the SLEUTH model. Before a prediction is made, the model needs to be calibrated, meaning that the optimal set of coefficients that describe the growth profile of the region needs to be found. This is

[2] 'Diffusion', 'Breed', 'Spread', 'Slope' and 'Road-gravity' starting with a capital always refers to the coefficient of the model. This applies to their abbreviations (Diff, Brd, Sprd, Slp and RG) as well.

[3] Pixel is the base unit on which the model simulates urban growth. The real-life extent of the area represented by a pixel depends on the spatial resolution chosen.

done by running simulations on historical data with different coefficient sets, comparing the results, and selecting the one with the best fit for the prediction.

The exact way of calibration has been at the forefront of SLEUTH-related studies since the model's inception (Dietzel & Clarke, 2007). The initial version applied monolooping – running the simulation from 1 to 100 for one of the coefficients while keeping the others constant, selecting the value producing the best fit, and repeating the process for the remaining 4 coefficients as well (Clarke, 2018). As this method was both computationally inefficient and very susceptible to local maxima, the next version used a brute-force method in which the model explored finer and finer ranges of the coefficient space in multiple phases. In a typical application the first phase of 'coarse calibration' explores the coefficient ranges in a 25-step interval, meaning that all permutations of (1, 25, 50, 75, 100) for all the five coefficients are tested. For the next round of calibration, the finer ranges around the best coefficients are explored with smaller steps, and the process is repeated until the best set of coefficients is found.

The multi-phase brute-force method remains the dominant calibration method for the majority of studies applying SLEUTH even today. Advances in computational power made the model more accessible by reducing the time needed for calibration, even though CPU power remains a common bottleneck for larger projects, and calibration times measured in days or weeks are still not unheard of (Clarke, 2018). Apart from its still high computational requirement, another drawback of the brute force method is its reliance on discretionary choices made by the modeler. Brute force calibration is a method in a broader sense and not an exact process: the authors of the model provide a guideline, but its implementation – number of phases, size of steps, selection criteria for the best coefficients, etc. – depends on the modeler. This led to divergent calibration strategies across studies, significantly undermining their comparability.

One of the latest advancements regarding model calibration is the application of Genetic Algorithms. GA aims to simulate biological evolution and natural selection to find the optimal solution. A set of 5 coefficients make up an individual gene, and the total population of the genes makes up the genome. The algorithm mimics evolution by running simulations with the genes, evaluating their performance, and creating a new generation of genes with a better fitness (better spatial fit). This is achieved by crossover – combining existing genes – where the parents are picked by tournament selection. To increase diversity, the genes undergo mutation, meaning that some coefficients of the gene are swapped with a random number. Survival selection is based on continuously replacing the weakest genes of the old generation with the strongest of the new one. This method is showed to maintain or even slightly improve the fitness or results compared to the manual calibration method, while vastly reducing the calibration time, up to by a factor of five (Clarke-Lauer & Clarke, 2011). GA requires eight additional hyperparameters to be defined such as mutation rate or genome size, but the newest version of the SLEUTH model includes an optimal setting for these by default, which has been derived after extensive research by the model's author and is aimed to be generally applicable across projects. (Clarke, 2018).

Regardless of the exact method, the calibration phase is a critical part of the modeling exercise as the results achieved with the consecutive prediction depend on how the model was calibrated (Silva & Clarke, 2002). With SLEUTH the calibration is achieved by backcasting – simulating urban growth for the past, comparing it to the reference data, and selecting the best coefficients based on the prediction's goodness of fit to the actual data. An integral part of the backcasting process is to define the 'goodness of fit', the evaluation criteria for the comparison of results achieved with the competing coefficient sets during calibration. The model itself calculates 12 different metrics measuring spatial fit for each calibration run using the Urban Growth Model, but the decision on how to use them to find the best spatial fit is left to the modeler (Dietzel & Clarke, 2007). Despite the model's long history, there is no clear consensus on how to apply these metrics for selection: some studies used only one metric (Silva & Clarke, 2002), some multiple ones (Jantz et al, 2004), and some used all of the available metrics provided by the model (Yang & Lo, 2003).

Figure 2

Built-in metrics of SLEUTH to evaluate the goodness of spatial fit

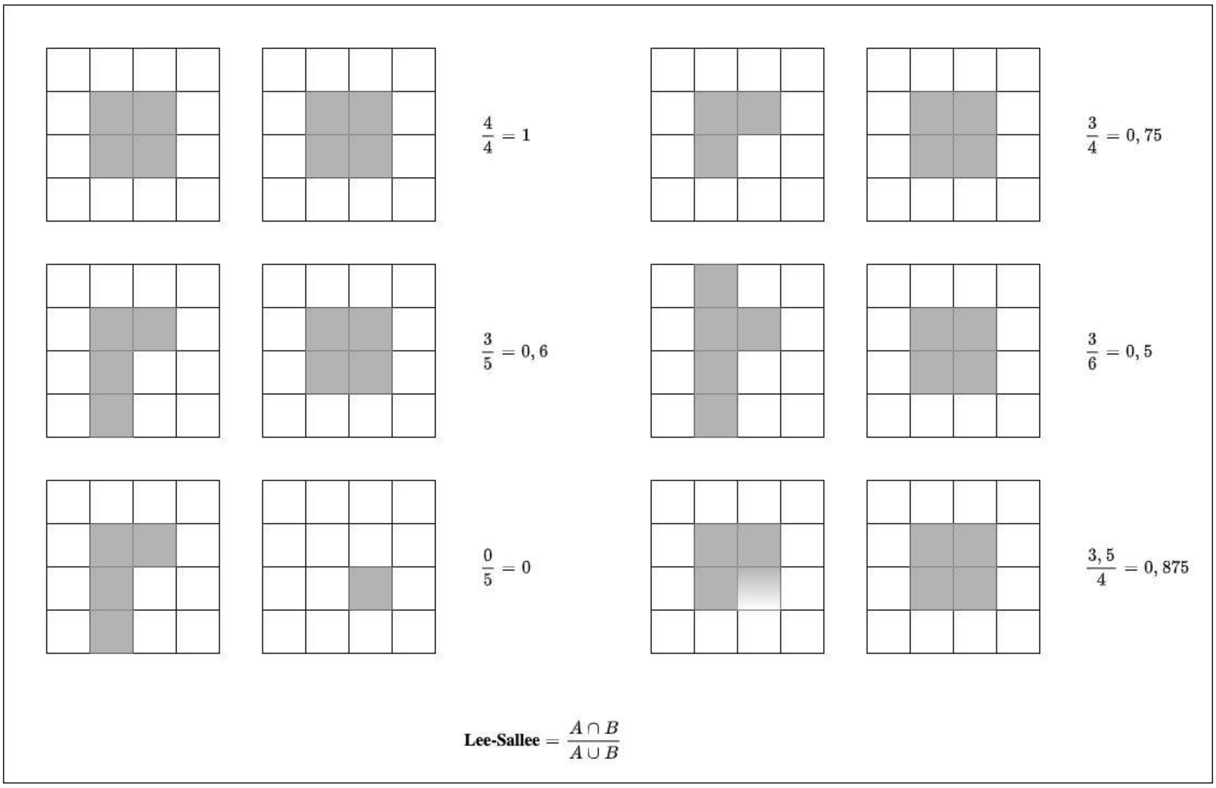
Metric Name	Description
Product	All other scores multiplied together
Compare	Modeled population for final year/actual population for final year, or IF $P_{modeled} > P_{actual}$ {1 – (modeled population for final year/actual population for final year)}.
Pop	Least squares regression score for modeled urbanization compared to actual urbanization for the control years
Edges	Least squares regression score for modeled urban edge count compared to actual urban edge count for the control years
Clusters	Least squares regression score for modeled urban clustering compared to known urban clustering for the control years
Cluster Size	Least squares regression score for modeled average urban cluster size compared to known average urban cluster size for the control years
Lee-Salle	A shape index, a measurement of spatial fit between the model's growth and the known urban extent for the control years
Slope	Least squares regression of average slope for modeled urbanized cells compared to average slope of known urban cells for the control years
% Urban	Least squares regression of percent of available pixels urbanized compared to the urbanized pixels for the control years
X-Mean	Least squares regression of average x_values for modeled urbanized cells compared to average x_values of known urban cells for the control years
Y-Mean	Least squares regression of average y_values for modeled urbanized cells compared to average y_values of known urban cells for the control years
Rad	Least squares regression of standard radius of the urban distribution, i.e. normalized standard deviation in x and y
F-Match	A proportion of goodness of fit across landuse classes. {#_modeled_LU correct/(#_modeled_LU correct + #_modeled_LU wrong)}

Note. OSM is composed of the compare, population, edges, clusters, slope, X-mean and Y-mean metrics. From: Dietzel, C., & Clarke, K. C. (2007). Toward optimal calibration of the SLEUTH land use change model. *Transactions in GIS*, *11*(1), p. 33 (https://doi.org/10.1111/j.1467-9671.2007.01031.x). Copyright 2007 The Authors.

Among the 12 metrics, the Lee-Sallee shape index (Lee & Sallee, 1970) gained prominence as the one often used alone: it measures spatial similarity between the calibration outcome and the reference data by calculating the ratio of the intersection and union of the two maps (Dietzel & Clarke, 2007).

Figure 3

Examples for the calculation of the Lee-Sallee metric for different spatial overlay scenarios



Clarke published their study intending to provide a unified metric to be used for later studies. Acknowledging that including multiple metrics might make the results more robust, while some of the metrics produce very similar outcomes and are therefore a source of possible bias, they set out to define an optimal set of metrics. They achieved this by creating a large amount of reference data for comparison and applying an unsupervised learning technique called Self-Organizing Map (Konohen, 1990) to cluster the metrics and find a subset of them to be used for a composite benchmark. This resulted in the creation of OSM – Optimal SLEUTH Metric – a product of 7 different spatial fit metrics. OSM saw widespread adoption after 2007, and also a calibration guide based on OSM was added to the Gigalopolis Project website – in addition to the already existing, Lee-Sallee metric – thus making it the de-facto official calibration method to follow.

Experiment setup

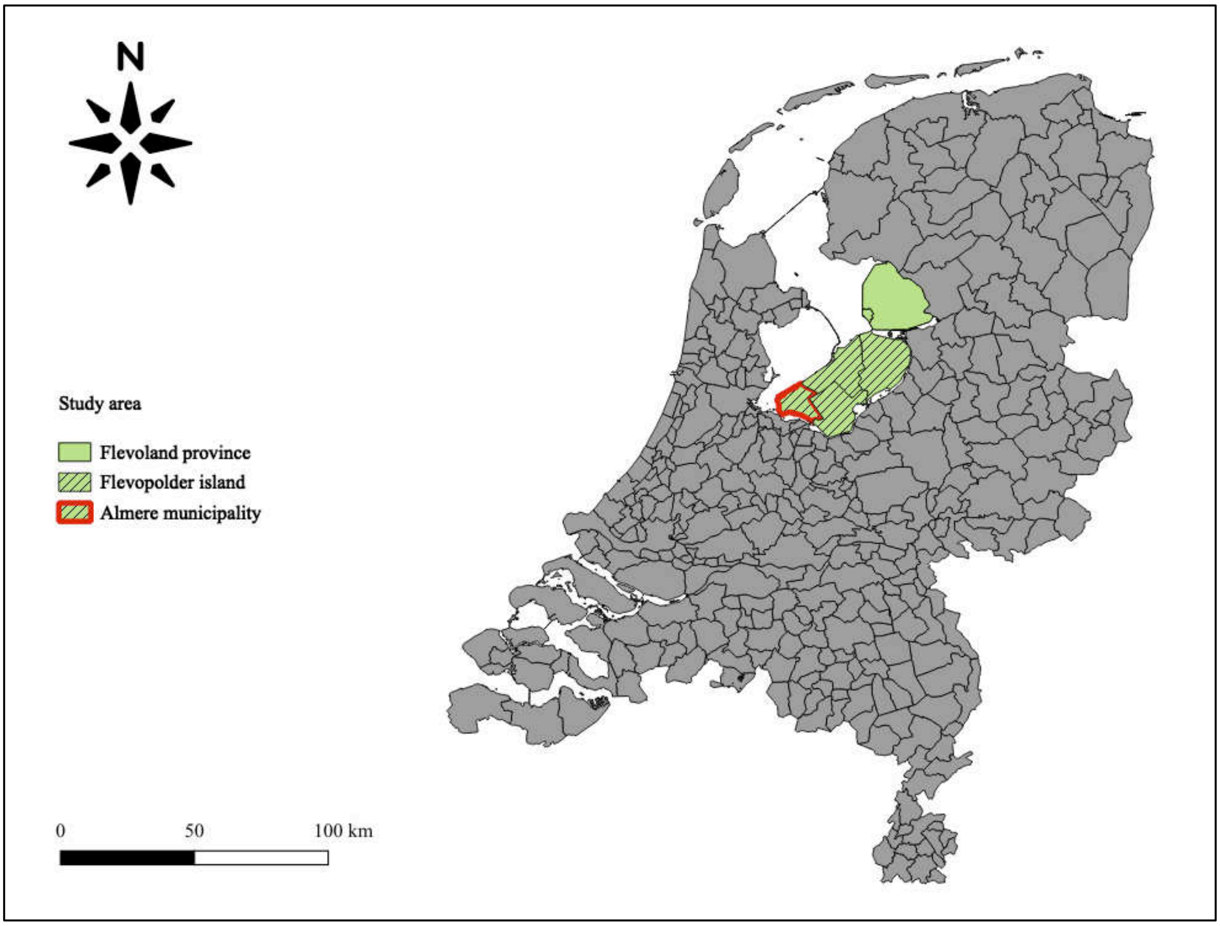
Choosing a city to model

The Dutch city of Almere has been chosen as a study area. This was partly due to the richness of available geographic data for the Netherlands, but was also partially motivated by its unique development history. Almere is a planned city – the island polder it was founded on – Flevopolder

- was reclaimed from the sea from 1959 to 1968, thus making it the youngest Dutch city. Currently it has the largest population in Flevoland municipality with a population of just above 200.000 citizens (Gemeente Almere, 2021), with plans to further expand its population to 350.000 by 2035 (Duivesteijn, 2007).

Figure 4

Location of the study area within the Netherlands



Although SLEUTH is designed to be general in the sense that it claims to be able to model different drivers of urban development through the growth coefficients, it has seen the majority of applications within the United States and in developing countries (Clarke, 2008). Its initial success in modeling the urban sprawl of typical American metropolises led to a wider adoption around the globe, especially within the developing countries of Asia, Africa, and South America. As of 2021, after the United States, the most cities modeled with SLEUTH are within China, Iran, and India[4].

[4] The research project that was the forerunner of this thesis reviewed 52 studies completed with SLEUTH. Among these, 8 modeled regions were in China (Hua et al., 2014; Liu, et al., 2012; Liu et al., 2019; Tan et al., 2009; Xi et al., 2009; Xibao et al.; Xu, et al., 2006; Yin et al., 2008), 5 in Iran (Bihamta et al., 2014; Dadashpoor & Nateghi, 2017;

The cities modeled in these countries typically went through rapid urbanization during the 20th century, without much central restriction on land use or strongly enforced zoning. This has resulted in unrestricted, organic growth patterns for these cities (often with informal settlements within the peri-urban zone) and cellular automata-based models like SLEUTH proved to be a fitting choice to represent this kind of development.

On the other hand, Northern and Western Europe remain underrepresented on the global map of SLEUTH studies: neither the United Kingdom nor France, Germany, or Sweden has seen any application so far. In the Netherlands, the only city that has been modeled with SLEUTH is Groningen (Venema, 2016). It is worth noting that compared to the developing world, Northern and Western Europe has been already significantly urbanized a century ago, while the growth rate of the population remained comparatively low. This resulted in less spatial growth for the cities of this region, thus making them naturally less attractive for this kind of modeling (Ferdinand, 2020).

Another attribute of this region – especially of the Netherlands' – is the relatively high level of urban planning. Although some studies confirmed the model's ability to accommodate for such growth patterns as well (Clarke et al., 2007), knowing cellular automaton-based models natural fit for organic patterns, it is reasonable to question if SLEUTH can also predict the strongly policy-driven development trends of this region with the same accuracy. It must be noted that testing this conjecture is not among the explicit aims of this study, but conducting a study on a Western European planned city fills a hole within the currently existing body of SLEUTH literature. This also means that – if SLEUTH will indeed have trouble picking up the patterns of policy-driven urban development – we can reasonably expect a lower spatial fit for the predicted outcome of the model. However, this is less of an issue for this current study, as the focus is on the comparative evaluation of different outcomes, and not on producing an actual 'best' prediction for the future of Almere.

Choosing a time period for study

It is not unusual to use data for calibration and prediction with SLEUTH that encompasses a period of more than half a century (examples include Clarke et al., 1997; Sangawongse et al., 2005; Abd-Allah & Mohamed, 2007), but the wish to conduct a study on a wide temporal horizon is often hampered by the lack of available historical data. The model requires multiple historical

Dezhkam et al., 2014; Rafiee et al., 2009; Sakieh et al., 2014) and 4 in India (Gandhi & Suresh, 2012; KantaKumar et al., 2011; Mahiny & Gholamalifard, 2011; Saxena & Jat, 2019).

versions for at least the urban extent, and finding such data for regular historical intervals can be a very resource-intensive task. Even though the model does not impose a limit on the length of period being modeled, extensive studies into the temporal sensitivity of the model have shown that SLEUTH produces superior results when used with shorter histories and shorter forecast horizons, and a long history usually produces a less convincing short-term forecast than a short one (Candau, 2002).

Due to limited resources available for this study, the selection of temporal scope was primarily driven by the availability of historical data. SLEUTH requires at least four unique urban extent control layers for adequate calibrations. This study encompasses two decades onward from 1996, with eight control layers instead - for 1996, 2000, 2003, 2006, 2008, 2010, 2012 and 2015. Thanks to the publicly available, good quality geodata provided by CBS, this study managed to model the urbanization process for Almere on a scale that is in line with expert recommendations (Candau, 2002; Dietzel & Clarke, 2004).

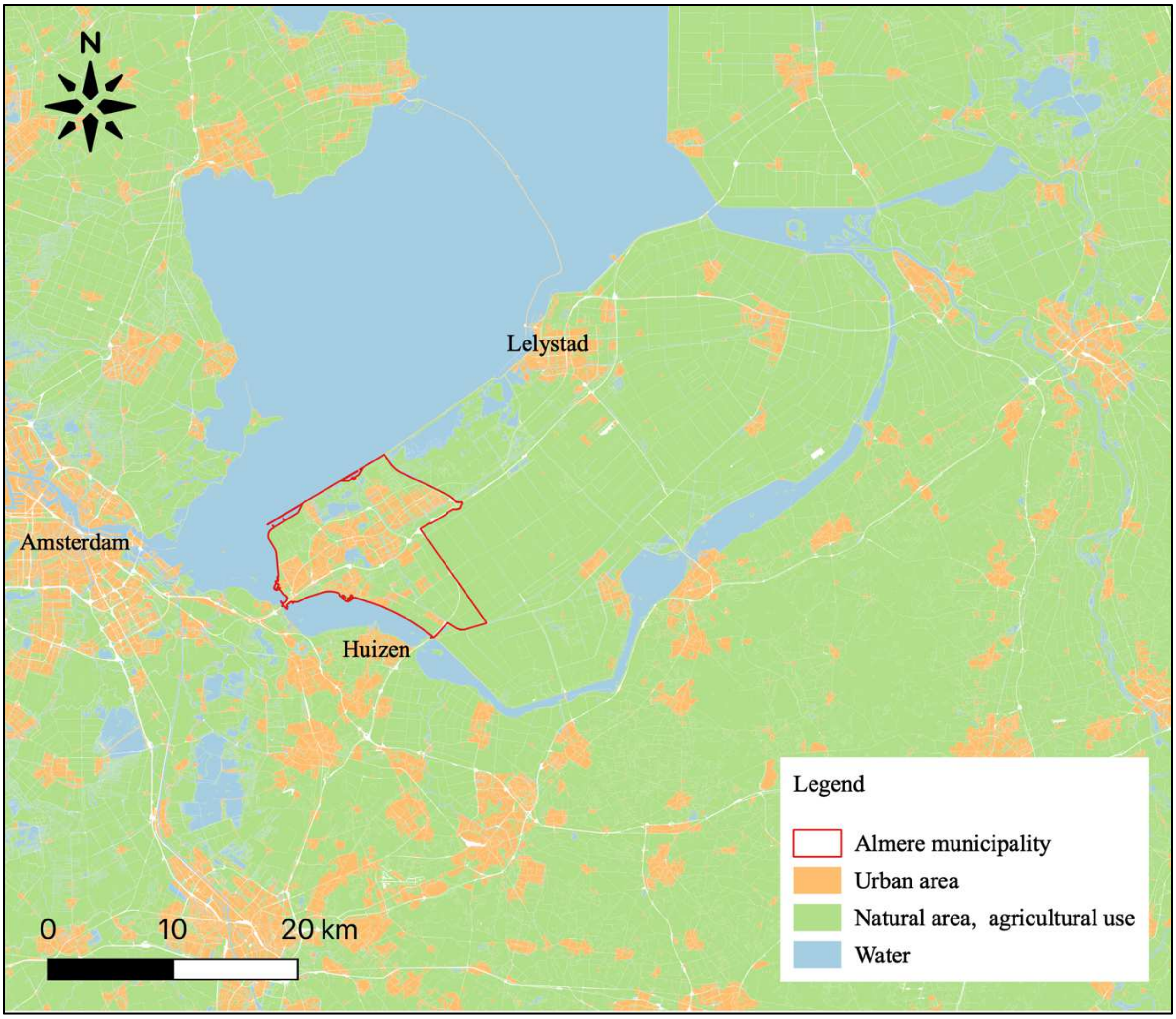
Choosing a spatial extent

As this study uses the UGM model to simulate land use change from non-urban to urban only, four different layer types – Slope, Exclusion, Urban and Transportation - are needed as input data. A fifth layer – Hillshade – is only used to provide a visual backdrop for the representation of urbanization process, but it does not affect the underlying calculations. All input layers need to be provided as grayscale GIF files with the same size, meaning that all input layers need to be rasterized while having the same resolution (pixel size), the same extent, and for valid results the raster layers must be perfectly aligned as well.

Given that this study aimed to prepare the input data sets for multiple extents, a decision needed to be made early on the exact map sizes. At the birth of the model, there was no unified approach or theory to address scale dependencies and make optimal scale choices (Obeysekera & Rutchey, 1997), and this issue remained unaddressed since (Jantz and Goetz, 2005). This deficiency was the main inspiration for the paper, but it also meant that map sizes had to be picked by discretionary choice, using only the natural setting of the study area as guidance. Almere lays on the western tip of Flevopolder island, therefore there is a natural limit to the extent of urban growth: the city is bounded by the sea from the west, has a roughly 10 km long north-south extent from coast to coast, and large portions of land available for further development are only available to the east. Almere city (Almere Stad) has a spatial extent of roughly 7 km × 7 km, but it is fused together with other settlements (Almere Poort, Almere Haven, Almere Buiten), thus indicating a natural smallest extent for meaningful modeling at around 10km[5]. The closest city on the island – Lelystad – is around 24km away (city center to city center), but within the same distance, there are comparable settlements on the other side of the inland sea, most notably Huizen. This study considers these cities as their own engines of urbanization, and to distinguish between their influence, this distance hints at a meaningful maximum modeling extent. To make meaningful comparisons, four spatial extents were picked within this range: 10km, 14km, 18km, and 22km. They are all centered around the same location (city center of Almere Stad) to facilitate comparison. The symmetric, 4km increase for the consecutive extents was another discretionary choice, aimed to simplify the comparative analysis later between core and periphery regions.

Figure 5

Urbanization around the greater study area (2015)

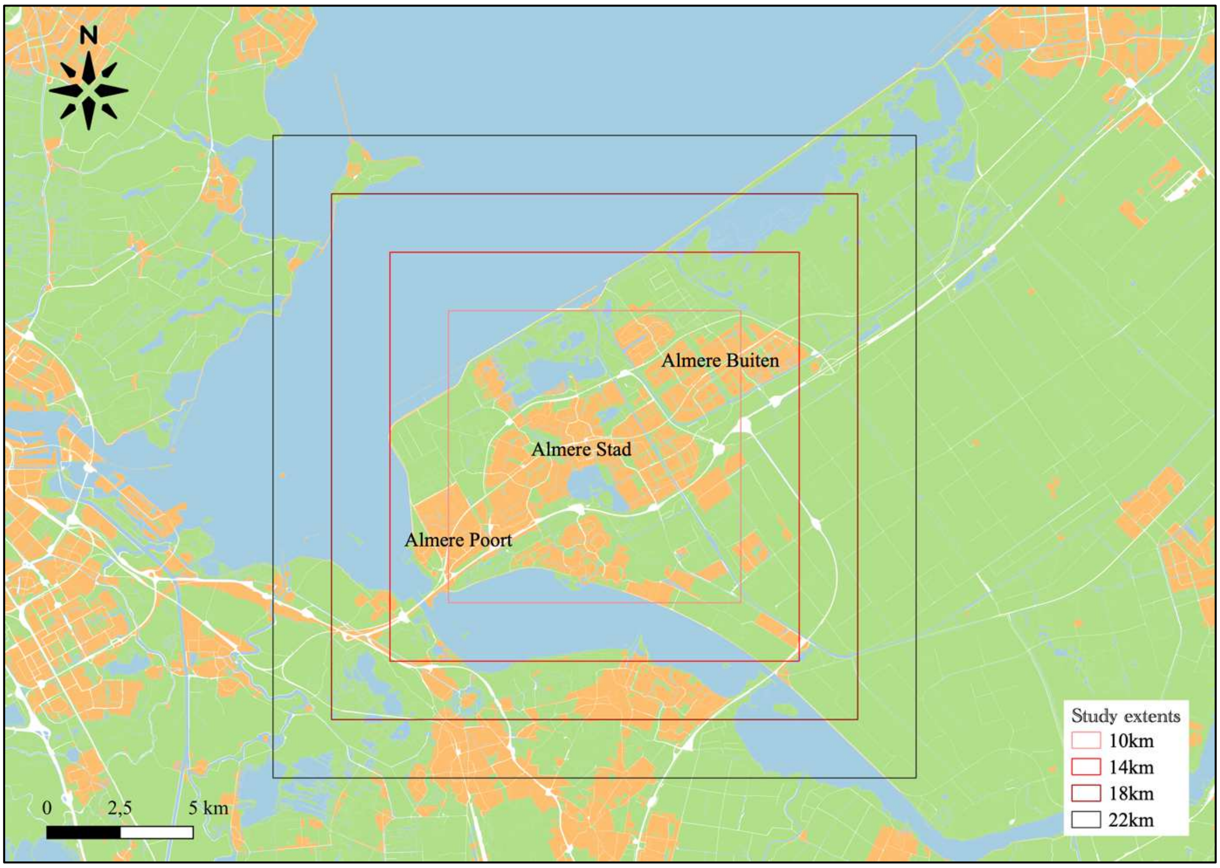


[5] SLEUTH requires input images with the same width and height, therefore all different extents used during the modeling exercise will have the same size in both directions. Extents will be referred to in the format of '10km', indicating a 10 km × 10 km area.

Due to the input layer requirements, a project resolution for all layers applied needed to be decided in advance. Modeling land-use change on a fine level is often obstructed by either the lack of available high-resolution data or by computational restrictions, but none of these factors impeded the current study. The source datasets were available either as vector or as 25m raster layers, thus supporting a wide range of options regarding resolution. Due to the location chosen for the study, computational requirements were also less prohibitive: even the largest selected extent (22km) was comparatively small, translating to quicker model execution. This enabled a resolution choice unhindered by computational concerns. Dietzel and Clarke (2004) – while highlighting that currently selecting a proper resolution is more of an art than science – provide a guideline for picking an optimal resolution based on map size, with a recommended resolution between 10-30m for small cities, defined by a spatial extent of less than around 32km. In line with this recommendation, a project resolution of 25m was chosen.

Figure 6

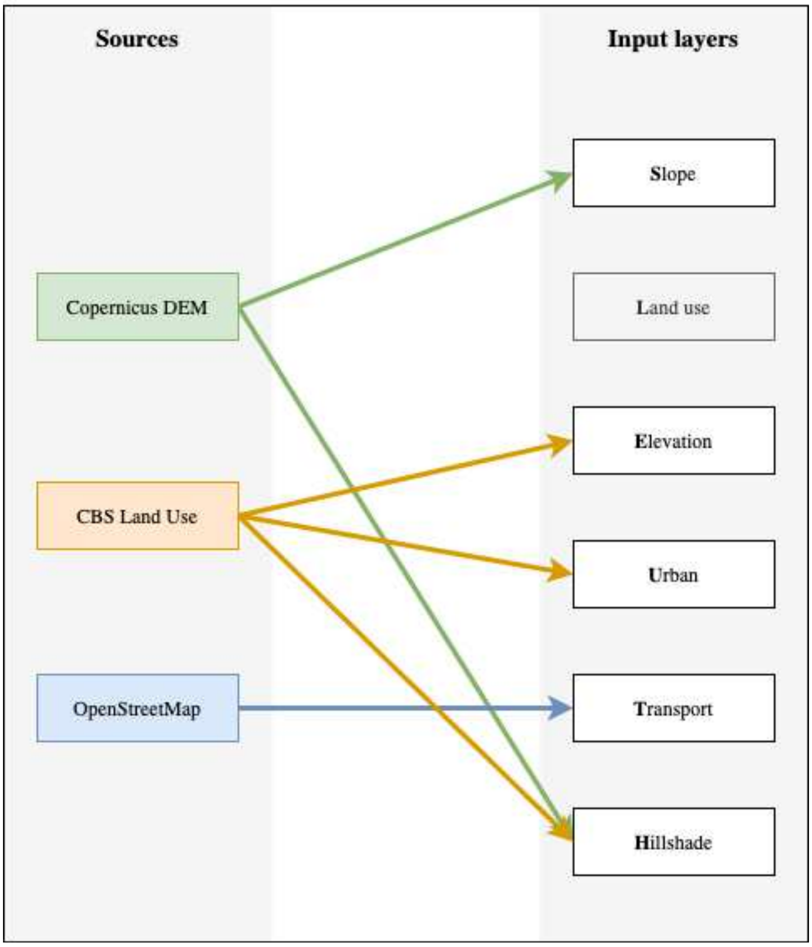
Comparison of study extents around Almere



The study used three data sources for the creation of input layers. Land use data for the Netherlands have been provided by CBS in a vector format. CBS creates and compiles land-use vector maps in roughly 3-year intervals, while also composing mutation maps where error corrections added to later versions are retroactively applied for earlier versions as well, to facilitate methodically sound multi-year comparisons. This study uses the mutation map set, which was used for the creation of the Exclusion, Urban and Hillshade layers. For the Transit layer, OpenStreetMap transportation data has been downloaded for the Netherlands in vector format. The model makes it possible to use multiple historical transportation maps, but given the short modeling period of fewer than two decades this was not deemed necessary, and only a single layer – representing data as of 2014 – has been used. Lastly, the Copernicus EU Earth Observation Program's 25m resolution Digital Elevation Model (DEM) maps were used for topographic components of the Hillshade and Slope layers.

Figure 7

Data sources used for the different input layers



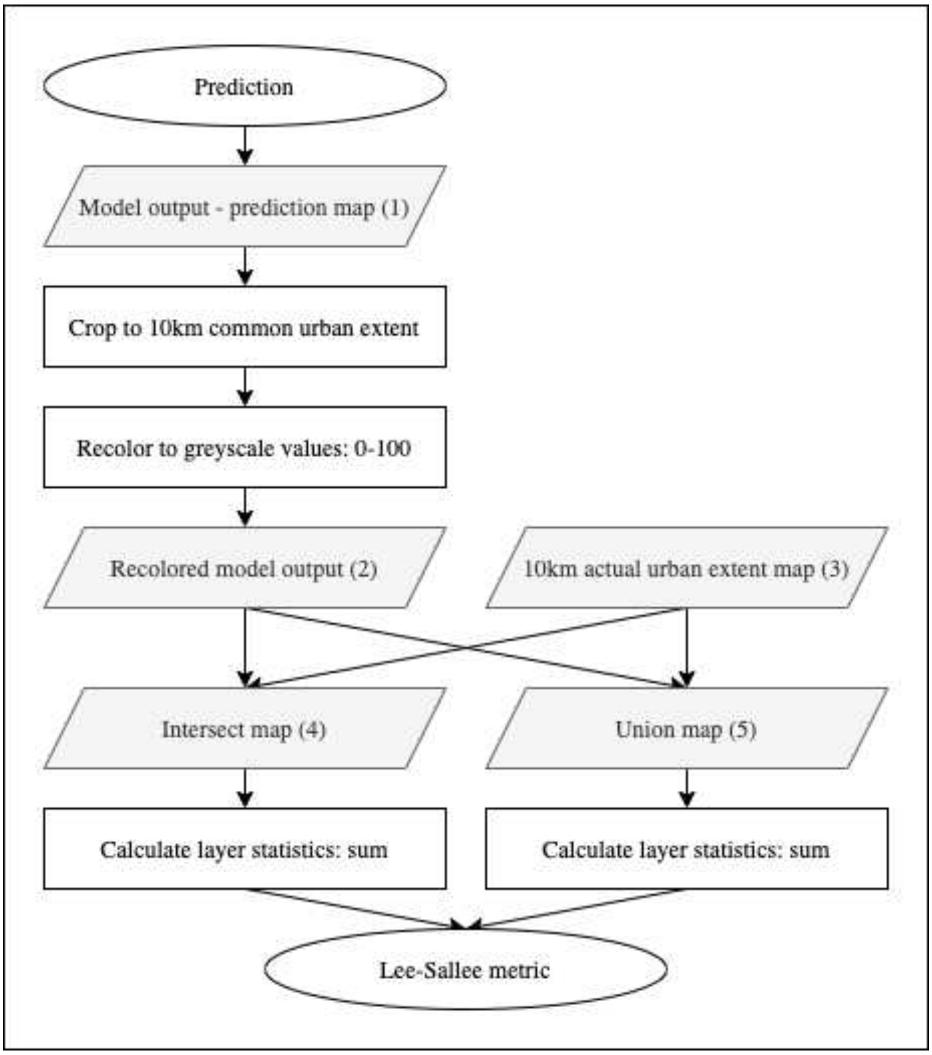
Measuring spatial fit

For coefficient selection during the calibration phase, this study relies on Optimal SLEUTH Metric. The motivation to use OSM was twofold. On one hand, it has favorable characteristics, namely that it proved to be sufficiently robust without including redundant component metrics that would risk including bias (Dietzel & Clarke, 2007). On the other, it is worth noting that its current status as the de facto official metric to use is a strong argument for its further use as well. The literature of SLEUTH significantly suffers from the lack of comparability among studies, which is due to the divergent methodologies being utilized, thus rendering the results hardly comparable across studies. This chiefly arises from the lack of standards – OSM mitigates this issue concerning the measurement of spatial fit. This study aims to propagate comparability by adopting this standard.

The OSM values can be acquired by running an auxiliary application after the calibration of the model is finished. This is convenient for comparing the overall spatial fit of multiple spatial extents, but cannot be applied for the comparison of the intersects of the different extents. This is because when comparing the intersects after the predictions are run for the 14km, 18km and 22km extents, their central 10km extents need to be cropped for further comparison, but SLEUTH does not provide tools for such analyses. For this comparison, the Lee-Sallee metric was used, which proved to be the most popular single (non-composite) spatial fit metric, with multiple studies (Silva & Clarke, 2002; Liu & Liu, 2009; Tan et al., 2009; etc.) relying solely on it for the calibration phase as well. An advantage of this metric is that due to its relative simplicity, it is easy to modify it further. SLEUTH outputs a probability-based prediction, meaning that in the map of predicted urbanization for each cell the percentage chance of being urbanized is provided through color-coding. This feature was utilized so that the Lee-Sallee metric was calculated in a probabilistic manner, meaning that the predicted chance of urbanization was compared to the actual data. This provided a much more faithful interpretation of the predicted results than a comparison based solely on binary urban/non-urban values. The Lee-Sallee metrics were calculated by using GIS: the probability outputs were cropped to the common 10km extent and recolored to single-channel values for easier comparison. Union and intersect maps were created between the recolored prediction results and the actual data, and the layer statistics values have been utilized to calculate the Lee-Sallee metrics for each extent.

Figure 8

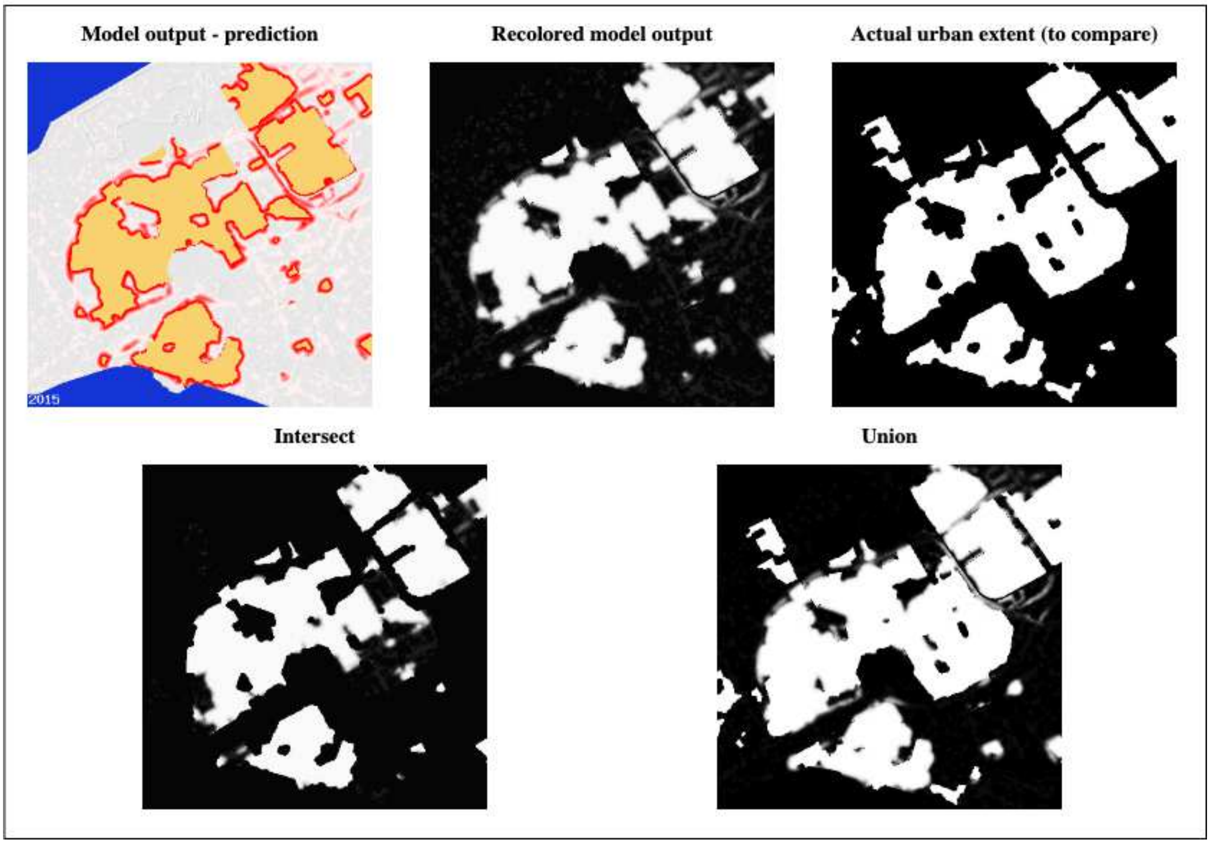
Process description for the calculation of the Lee-Sallee metric



Note. Examples of the artifacts used during the process (numbered) are shown in Figure 9.

Figure 9

Example artifacts during Lee-Sallee calculations



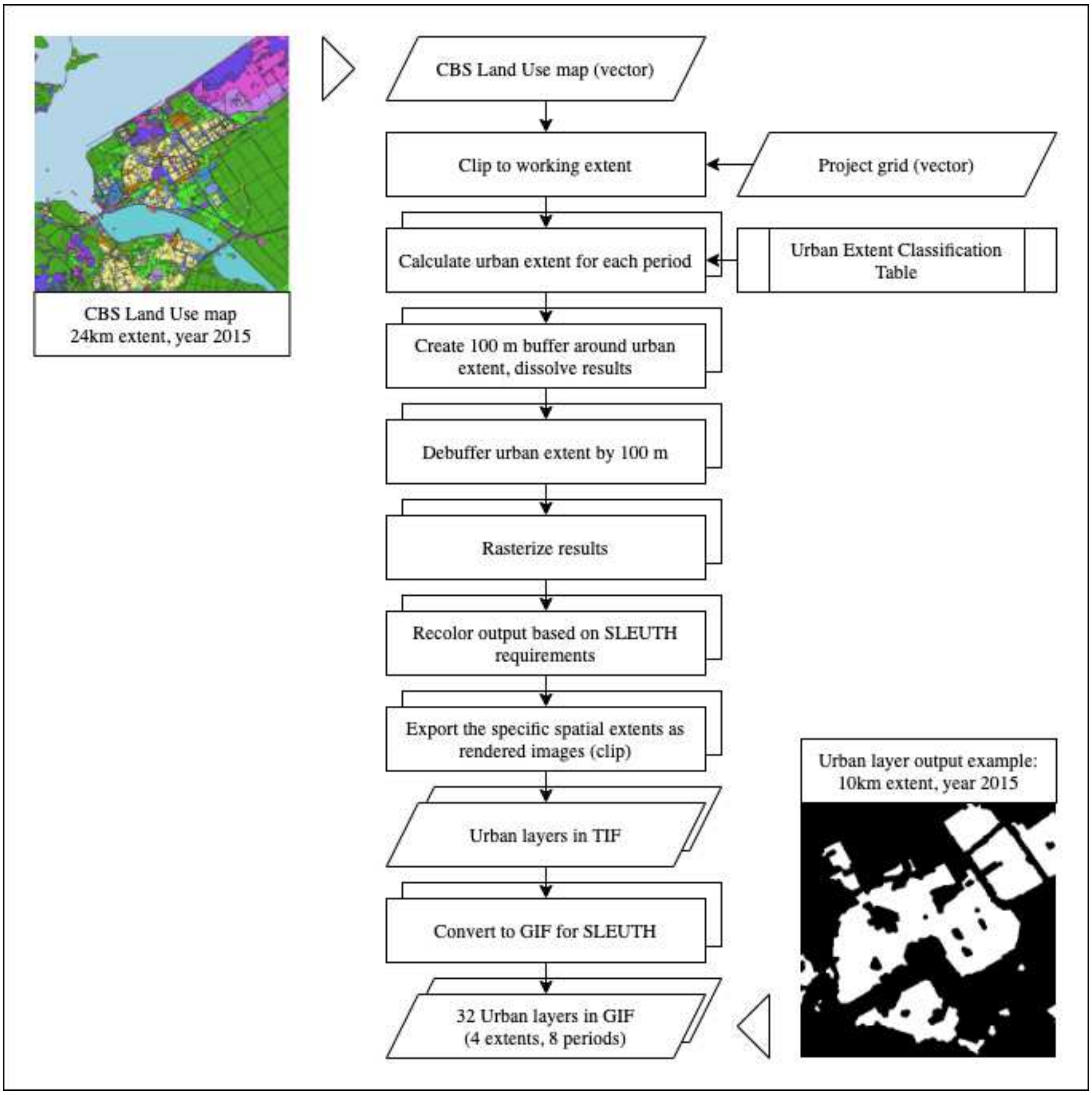
Data preparation

The first GIS-related technical challenge was to define a workflow that handles multiple spatial extents. To achieve this aim, a 1km grid system was created with a total working extent of 24km, where the manually defined origin was the city center of Almere. To speed up data transformation, all data sources were clipped to the working extent, and the subsequent transformations were applied to this extent as well. The differently sized input maps were created only at the end of the data preparation processes, by selecting the appropriate extents (10km, 14km, 18km, or 22km) based on the grid.

For the Urban layers, the vector land-use data has been reclassified to either urban or non-urban. To obtain a less noisy result, the urban extents were buffered and debuffered by 100 meters. The outputs were dissolved and rasterized.

Figure 10

Process description for the creation of Urban layers

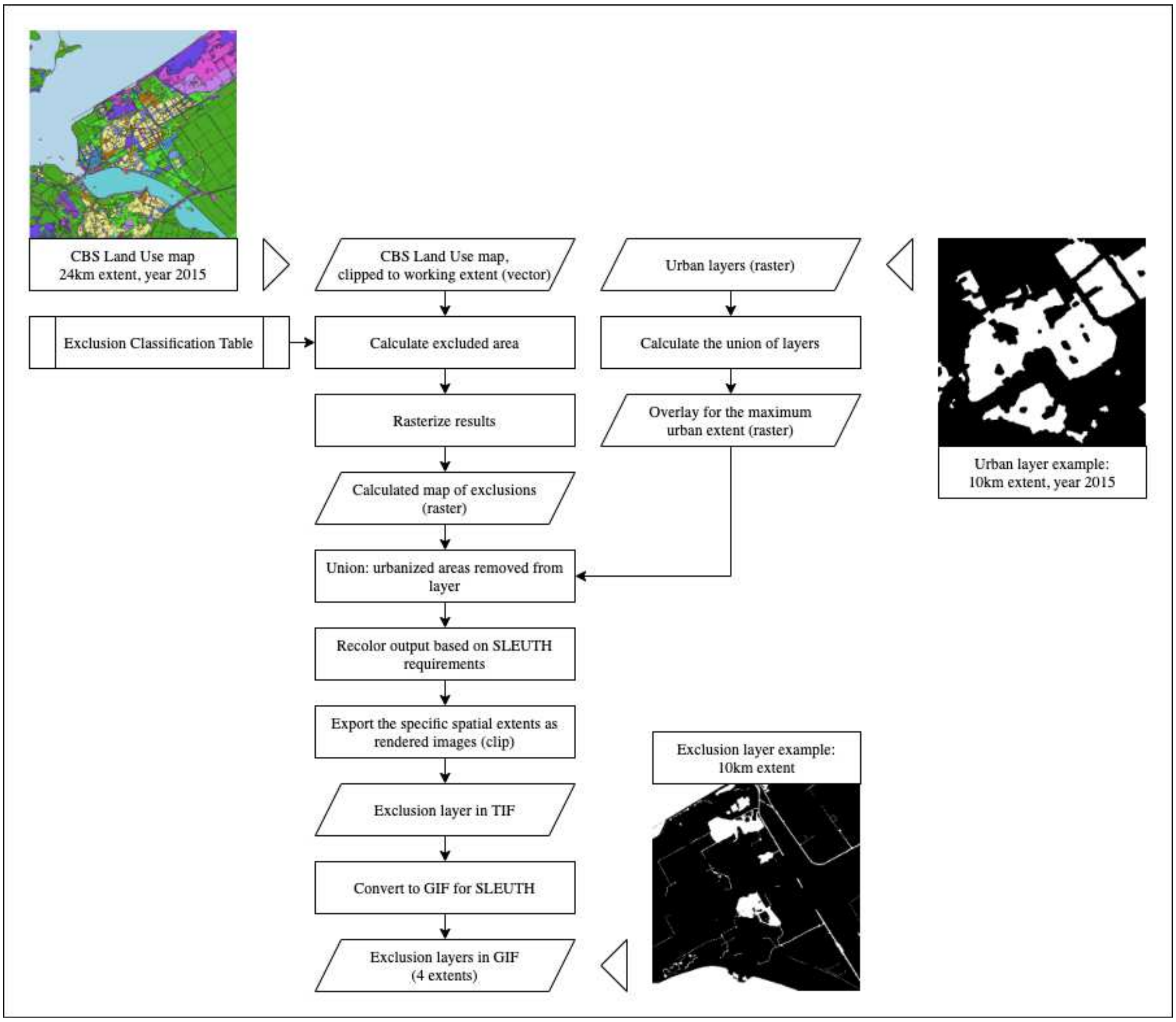


Note. See Table 6 (Appendix) for the Urban Extent Classification Table.

The Exclusion layer utilizes the same source file as the Urban layers. The land-use data has been reclassified to either available or excluded from urban development and rasterized. The model requires consistency in that area that will eventually become urbanized should be marked available on the Exclusion layer as well. To achieve this aim, a union of the eight Urban layers was calculated and overlaid – another union calculation – on the non-excluded extent.

Figure 11

Process description for the creation of Exclusion layers

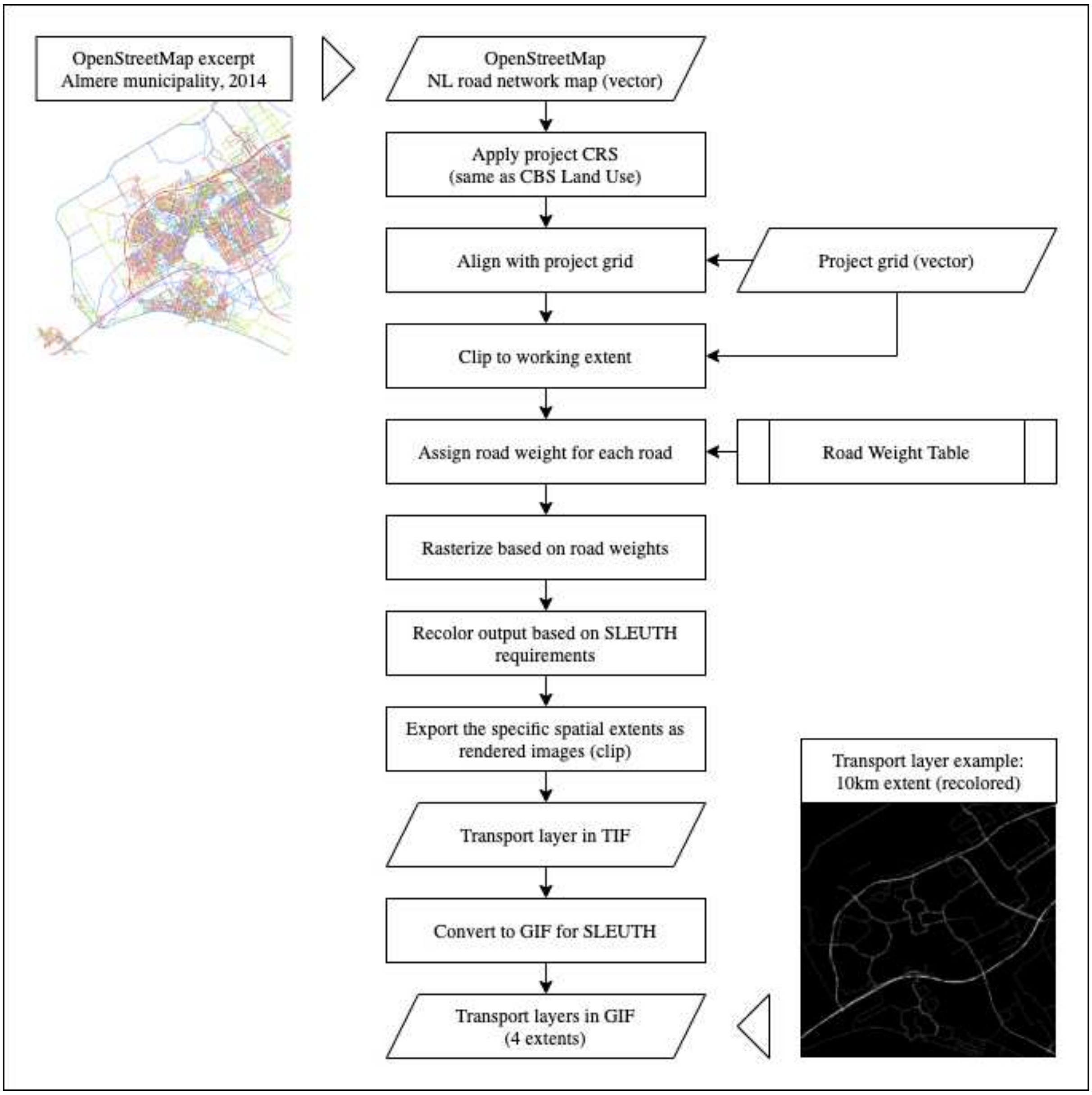


Note. See Table 7 (Appendix) for the Exclusion Classification Table.

The Transport layer allows for road weights to be defined, to account for the fact that larger roads must have an also larger effect on urbanization as well. Road weights were assigned to each road based on designation (motorway, primary, secondary, tertiary), and the layer was rasterized according to the road weights.

Figure 12

Process description for the creation of Transport layers

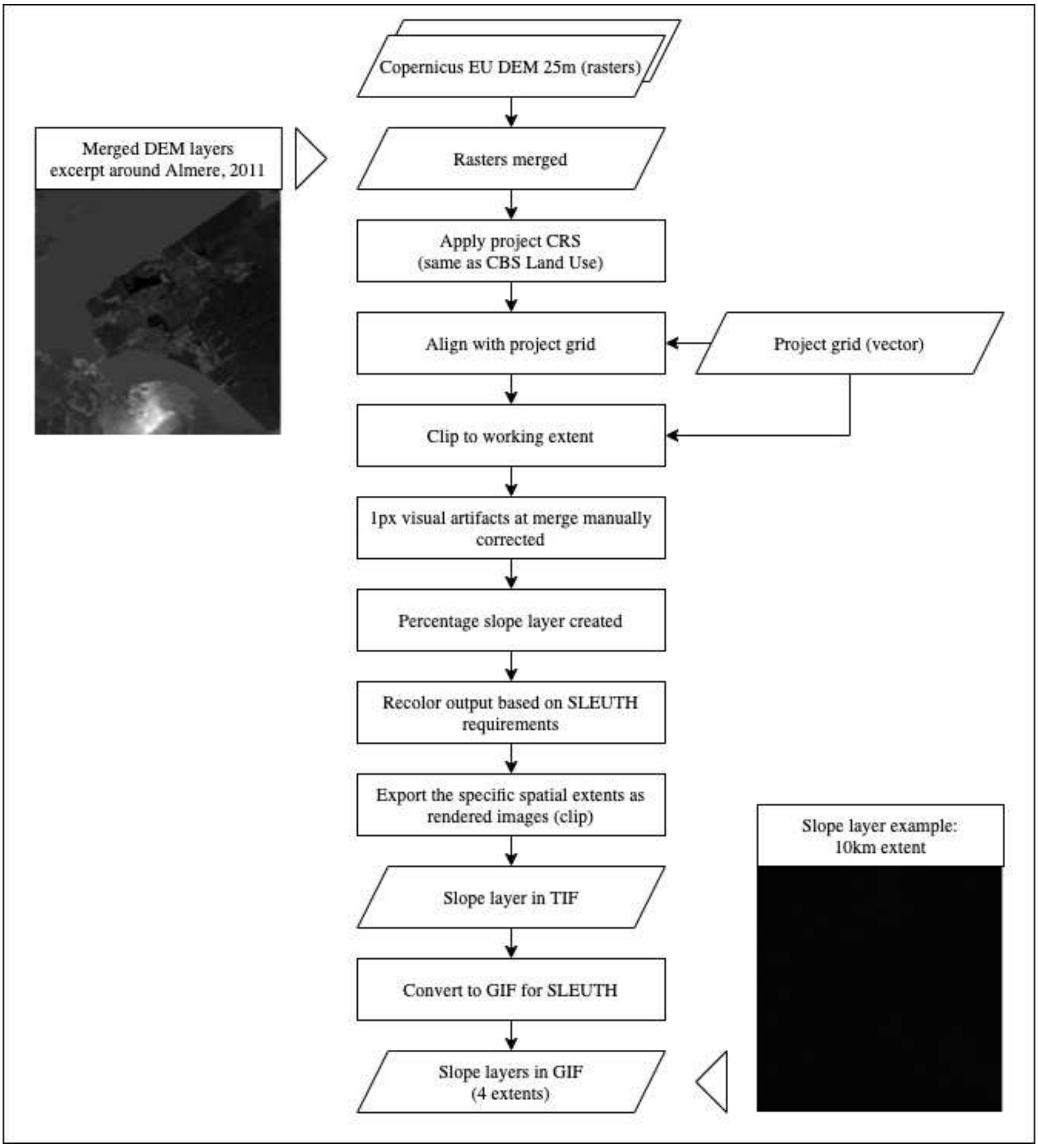


Note. See Table 9 (Appendix) for the Road Weight Table.

The Slope layer was the only input layer that used exclusively raster data as source, therefore no additional rasterization was necessary. The Copernicus EU DEM 25m dataset is sectioned to 1000km tiles covering the whole of Europe, but Almere lies on the border of two tiles, therefore it was necessary to merge them to cover the whole working extent of the project. This produced 1 pixel-wide visual artifacts at the joining, which were manually corrected by averaging based on neighboring values. To comply with SLEUTH's input requirements, the DEM model has been transformed into a percentage slope layer. It is worth noting that due to the flat terrain around Almere, the percentage slope layer is almost completely black.

Figure 13

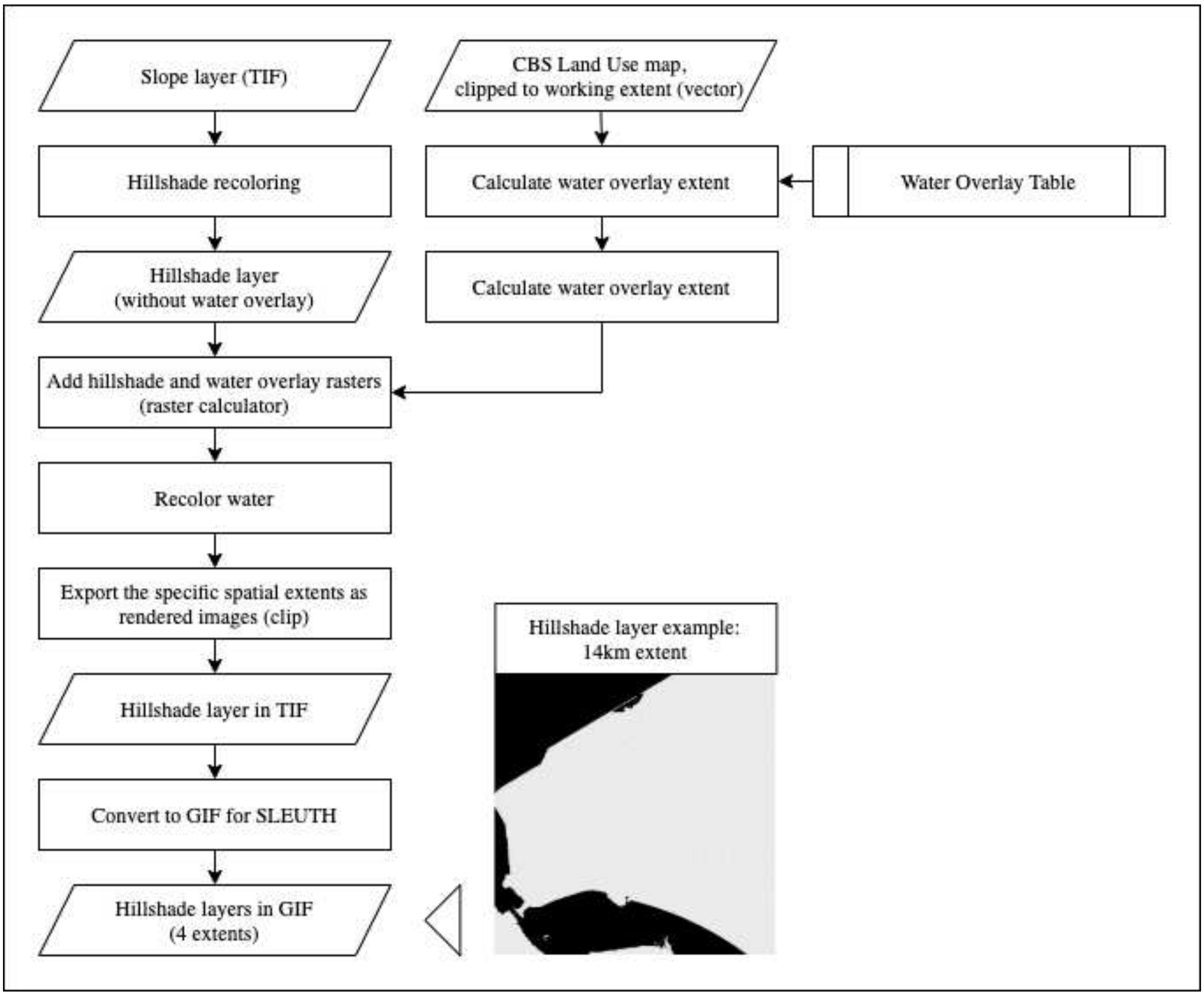
Process description for the creation of Slope layers



The Hillshade layer used the Slope layer as an input, with hillshade recoloring applied to provide more visual cues. Although it is not necessary, the model makes it possible to define a specific output color for open water to increase clarity. To achieve this aim, the land-use data has been reclassified to open water or non-open water and rasterized. The two raster layers together produced a Hillshade layer with the open water overlaid.

Figure 14

Process description for the creation of Hillshade layers



Note. See Table 8 (Appendix) for the Water Overlay Table.

Calibration phase

Due to the expected gains in computational speed and its non-reliance on user-made decisions, the GA method was the preferred mechanism for calibration, with the brute force method being used only as a benchmark and backup. Decreasing the effect of user-made choices on the calibration process, thus replacing the subjective decisions with an objective process is an important target to improve modeling quality (Sakieh et al., 2015; Jafarnezhad et al. 2016), and GA represents a significant improvement in this direction. Compared to brute force calibration, GA indeed achieved a substantial decrease in calibration time, while achieving similar levels of spatial fit as of the brute force method. The results, however, were disturbingly non-robust: simulations ran with the same parameters on the same input data produced vastly different growth coefficients. This was also the case when the number of Monte Carlo iterations was increased, thus the results of individual test runs were averaged over more instances. The table below shows a selection of the calibration results achieved with GA. The eight runs shared the same setup, GA hyperparameters, input data, and were run on the same hardware – the only variables changed were the number of Monte Carlo iterations (MC) and whether the self-modification function was enabled. Each scenario was run twice (banded rows). Even though the spatial fits achieved as measured by OSM were similar, the best individual coefficients diverged significantly, even among runs with the same setup.

Table 2

GA calibration results

Run ID	**MC**	**Self-mod**	**OSM**	**Diff**	**Brd**	**Sprd**	**Slp**	**RG**
6	1	YES	0.23616245	2	36	77	22	85
7	1	YES	0.23618980	48	50	9	65	62
8	4	YES	0.16068622	79	78	14	72	84
9	4	YES	0.19298603	10	21	41	21	21
10	1	NO	0.21115811	90	34	11	71	78
11	1	NO	0.19227983	35	32	36	31	39
12	8	NO	0.21237037	3	51	72	56	46
13	8	NO	0.19568364	3	94	46	15	87

Note. The table shows the OSM and coefficient results for the best performing genes per experiment only. The configuration files, outcomes and OSM statistics for each individual run are reported as in the Empirical Portfolio, and can be looked up by the Run ID.

The non-robustness might be less of an issue if prediction into the future is the aim of the modeling exercise and the spatial fit is still high. Different coefficient sets producing similar results can indicate that the area under study has no clear pattern in terms of spatial growth drivers, or that one or two growth drivers are insignificant, therefore their coefficients have little to no effect on the prediction outcome. This is the case for the slope coefficient for the current study as well: as the area of study is almost completely flat, this coefficient has little to no effect on the outcome, thus the high variance of the slope coefficient per calibration run is expected. But as the current study aims to compare the coefficients produced by different calibration runs as well, this poses a problem. Therefore, to guarantee more robust and comparable results, the model was recalibrated with the brute-force method.

One of the main drawbacks of the brute-force method is its reliance on user-made choices – in other words, the brain is replaced with brawn (Goldstein, 2004). Nevertheless, the effect of arbitrary modeling decisions can be minimized. This study follows the ‘baseline’ three-step calibration, which is the suggested calibration method outlined on the Gigalopolis project website, and it is also the calibration method used by the majority of studies applying SLEUTH. Even though this does not eliminate the need for user-made choices (especially for the coefficient selection method), this does increase the comparability with other studies. The other main drawback of the brute-force method is its high computational requirement, but this was less of an issue, given the relatively small spatial and temporal scope of the study. To safeguard against running out of computational resources and time, this study was run parallelized in a cloud environment. The calibration for the four different spatial extents to be compared can be run independently from each other, therefore multiple servers were prepared to be spun up on-demand, with the possibility to increase server performance if needed.

There are three areas where a deliberate decision was made to diverge from the baseline method. First, inspired by the research of Jantz et al., the self-modification function of the model was disabled. This was due to the lack of transparency into this functionality and how it affects the robustness of outcome, and also because during the relatively short modeling period we can reasonably anticipate a roughly linear development trend. (Jantz et al., 2003; Jantz & Goetz, 2005).

The second change relates to the map resolutions used in the subsequent calibration rounds. To save on calibration time, the guide recommends downsampling the input dataset to quarter resolution (25m to 100m) for the coarse calibration, downsample to half resolution (25m to 50m) for the fine calibration, and use the full resolution (25m) data only for the final calibration. However, the research of Dietzel and Clarke showed that using this hierarchical resolution model can lead to worse results – meaning that the coefficient set calculated will achieve a lower spatial fit, compared to using only the full resolution dataset (Dietzel & Clarke, 2003). On the other hand, given the relatively small spatial and temporal scope of the study, the computational time is inherently less of an issue for the study.

The third difference is that the coefficient selection process puts a bigger weight on the top results instead of the averaged results. This is due to the observation after running preliminary tests with the model that if the selection is based purely on observing the range around the average, the follow-up (second, third best) values can distort the average so much that the previous round's best value is not included in the new range – which can lead to the model converging on local maxima, and even achieving a worse spatial fit than in the previous round. To mitigate this issue, the coefficient selection always considers the range between the best and the average value, instead of simply exploring the range around the average result. The downside of this method is that due to the increased search range, the number of computations required to execute a phase can increase, but this was less of an issue for the actual project setup.

Following the guidelines outlined on the project website, the first step of the calibration (coarse calibration) explored the whole range of the coefficients with a step increment of 25 – meaning that the values of [1,25,50,75,100] were evaluated for all permutations for the five coefficients, bringing the total to 3125 combinations. These combinations have been run over 4 Monte Carlo iterations and averaged, for each year of the 2 decade-long calibration period [1996-2015]. This brings the total number of yearly simulations executed to 3125*4*20 = 250 000 for each spatial extent, meaning a total of 1 million simulations for the coarse round of calibration. Each of the 3125 runs averaged over 4 MC iterations produced one of the control statistics sets – the 12 metrics (plus OSM) that evaluate the spatial fit, representing a specific set of coefficients.

Even though the project website gives guidance about the coefficient selection for the following steps, in practice the remainder of the calibration process typically relies on expert decisions, as there is no exact guideline to follow. The lack of a common protocol does not only hurt inter-study comparability but would seriously undermine the comparability of the four different extents in this current study as well. To mitigate this issue, based on the Gigalopolis project guidelines a protocol was defined and followed for calibrating all the extents to ensure comparability. The following rules were defined:

- Calibration is done in three phases (coarse, fine, final)
- The full resolution is used for each calibration phase
- OSM is used for coefficient selection
- All results are averaged over 4 MC runs
- The self-modification mechanism (boom and bust cycles) is disabled

Complying with these rules, the following protocol was executed for each coefficient:

1. **Coefficient selection I – coarse calibration**
   1.1. Run the calibration: explore the whole coefficient range in 25-step increments [1,25,50,75,100]
   1.2. Select the best-performing single coefficient set ("best")
   1.3. Select the three best-performing coefficient sets and average them ("average"). If rounding is necessary, round away from the best result.

1.4. Explore the range between the best and average results. The best result should be included exactly, while the average should be inclusively in the range.

1.4.1. Optimally the range should be explored in steps of 5, corresponding to 7 steps provided that the range is 35.

1.4.2. If the range is larger, increase the step size to a maximum of 12. If the range is smaller, decrease the step size.

1.4.3. If the average equals the best prediction, which is an edge case (1 or 100) explore the half-range of the calibration step around it (1-12 or 88-100)

1.4.4. If the average equals the best prediction, which is mid-range, explore the half-range of the calibration step (12 in both directions) around the best prediction.

2. **Coefficient selection II – fine calibration**

2.1. Run the calibration: explore the coefficient range defined by the end of the coarse calibration phase

2.2. Select the best-performing single coefficient set

2.3. Select the three best-performing coefficient sets and average them. If rounding is necessary, round away from the best result.

2.4. Explore the range between the best and average results. The best result should be included exactly, while the average should be inclusively in the range.

2.4.1. Optimally the range should be explored in steps of 1, corresponding to 5 steps provided that the range is around 5.

2.4.2. If the range is larger, increase the step size to a maximum of 5.

2.4.3. If the average equals the best prediction, which is an edge case (1 or 100) explore the half-range of the calibration step around it.

2.4.4. If the average equals the best prediction, which is mid-range, explore the half-range of the calibration step around the best prediction.

3. **Coefficient selection III – final calibration**

3.1. Run the calibration: explore the coefficient range defined by the end of the fine calibration phase

3.2. Select the best-performing single coefficient set to be used for the prediction phase.

The table below shows the results of the calibration procedure – the columns correspond to the different spatial extents, while the three vertical sections represent the three phases of the calibration. Altogether 16 calibration runs were necessary to achieve a final set of coefficients to be used for prediction. The table lists the top performing coefficient set (in the usual order of Diff, Brd, Sprd, Slp, RG), the average of the top three coefficient sets, and the OSM value of the top performing coefficient set, which is expected to increase after each phase. For the final calibration, the Lee-Sallee metric for the best performing coefficient set is also reported. The selected ranges show the ranges chosen to be used for the next phase, in the format of [#start_value]-[#end_value] /[#step].

Table 3

Brute force calibration results

		10km	**14km**	**18km**	**22km**
Coarse	Run ID	101	102	103	104
	Best Result	1 100 50 25 1	1 100 100 100 75	1 1 100 75 1	1 1 100 100 1
	Top 3 AVG	1 75 34 50 34	1 50 34 75 58	1 1 100 75 17	1 1 83 83 42
	OSM	0.21231537	0.01387410	0.23147511	0.10093734
	Selected range	0-12 /3 75-100 /5 34-50 /4 25-50 /5 0-35 /7	0-12 /3 50-100 /10 34-100 /11 75-100 /5 55-75 /4	0-12 /3 0-12 /3 88-100 /3 63-87 /6 0-18 /6	0-12 /3 0-12 /3 80-100 /5 80-100 /5 0-42 /7
Fine	Run ID	111	112	113	114
	Best Result	1 75 50 25 35	1 70 89 90 55	1 3 88 63 18	1 1 95 85 28
	Top 3 AVG	1 92 50 25 14	1 70 89 90 59	1 2 96 71 12	1 4 97 85 16
	OSM	0.21375561	0.01522657	0.23493403	0.12851638
	Selected range	1-3 /1 72-92 /4 48-52 /2 23-27 /2 11-35 /4	0-3 /1 66-74 /2 83-95 /3 86-94 /2 55-59 /1	0-1 /1 2-3 /1 88-96 /2 63-71 /2 12-18 /2	0-3 /1 0-4 /1 95-97 /1 83-87 /1 16-28 /2
Final	**Run ID**	**121**	**122**	**123**	**124**
	Best Result	**2 84 50 27 23**	**1 70 89 90 55**	**1 3 96 67 12**	**1 1 97 86 20**
	OSM	**0.2182**	**0.0152**	**0.2910**	**0.1391**
	LS	**0.6821**	**0.6589**	**0.6866**	**0.7185**

Prediction phase

Results of the prediction runs

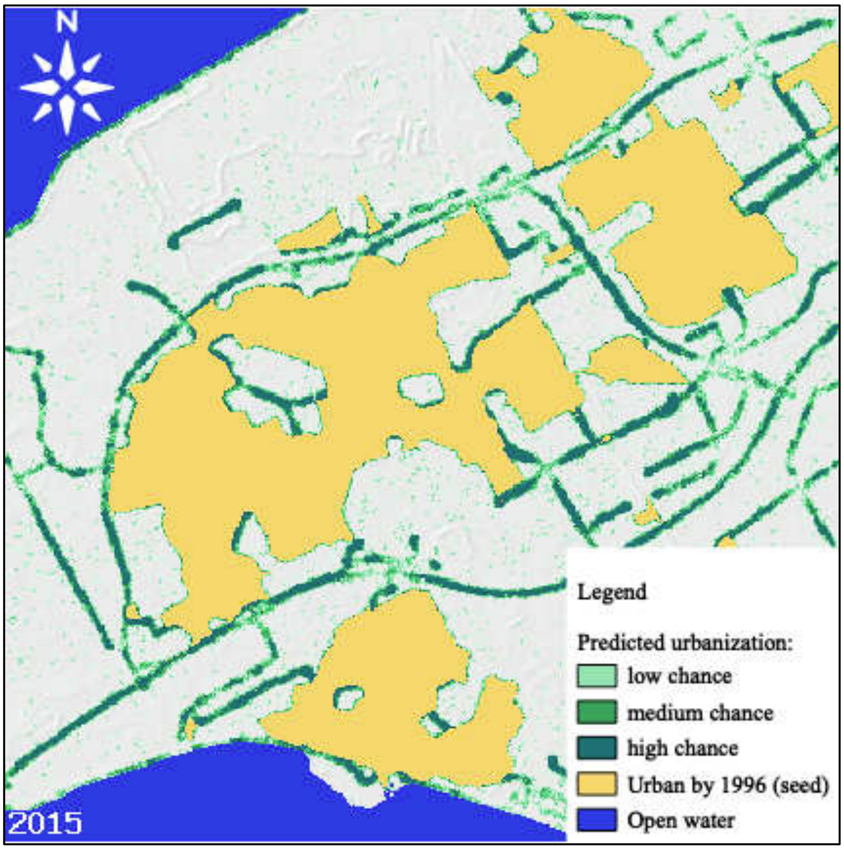
Compared to the calibration process, running a prediction is relatively straightforward, with less user interaction required. The best coefficient sets were chosen at the end of the final round of calibration and used for making the predictions. The seed year was chosen to be 1996 – the first year with data available – and the prediction was run until 2015 – the last year with control data. For measuring the spatial fit between the prediction and actual data for the intersection of extents

(10km central area), the year of comparison was 2015 as well. The seed year was chosen to be the earliest possible to maximize the expected variance between the extents, thus making the comparison easier. The closer the seed year is to the prediction year (for instance 2013 to 2015), the more accurate outcomes can be expected – but the extent of newly urbanized land within the shorter period will be smaller as well. If the prediction exercise aims to make as accurate future predictions as possible, it makes sense to select the latest possible year with data available for the seed year. But for this comparative analysis, to maximize the expected divergence between the different extents, the largest possible prediction period was defined. The predictions for each extent were run 100 times and averaged in a MC fashion. For the outcome, a probability color table was defined, which distinguishes between seed cells (already urbanized in 1996), cells with no predicted urbanization and cells with a given predicted probability of becoming urbanized until 2015. For the latter category decile ranges have been defined, representing a 1-10%, 11-20%, 21-30% etc. chance of becoming urbanized.

It is worth noting that during the research, both the calibration and the prediction phases needed to be repeated multiple times. The prediction outcomes were evaluated visually to confirm whether they 'make sense'. In some cases, when the literature provided no exact guidance on a specific topic, the investigation had to resort to trial and error methods. One typical example relates to the classification of different land-uses for the Exclusion layer. The first iteration of the model followed a more complex Exclusion classification, with differentiated categories: for instance, dry grasslands were categorized as more suitable than wetlands, and areas marked as building terrain were even more suitable for further development. This posed a problem though, as the highway-building projects are also classified as building terrain, therefore achieving a lower exclusion level (more suitable for urbanization) than the grasslands they were going through. This resulted in unrealistic spaghetti-like urbanization patterns along the highways, therefore the modeling exercise needed to be restarted with a simplified Exclusion layer. Nevertheless, it must be highlighted that even though such trial and error methods and visual confirmations are typical part of the modeling exercise, they represent human interference, thus possible adding further bias (Goldstein, 2004). Such decisions must be made in line with the integrity of the study, without propelling it towards a 'desirable' outcome.

Figure 15

’Spaghetti-like’ urbanization patterns predicted after misclassified Exclusion categories (10km)



Question 1 – differences in the coefficients

The first look on the coefficients after calibration shows that there are significant differences in the outcome, even between neighboring extent sizes like 10km or 14km. This challenges the idea of regional DNA proposed by Silva and Clarke (Silva & Clarke, 2002; Silva, 2004) – how can the coefficients describe the growth pattern of a region, if increasing the study area produces vastly different results?

Table 4

The coefficients after final calibration and their standard deviation

	Diff	Brd	Sprd	Slp	RG
10km	2	84	50	27	23
14km	1	70	89	90	55
18km	1	3	96	67	12
22km	1	1	97	86	20
STDEV	**0,43**	**37,83**	**19,30**	**24,94**	**16,38**

However, a more careful look yields more insights. The Diffusion coefficient has by far the lowest variance, as it has a minimal value for all extents. This coefficient is responsible for spontaneous growth, which represents urbanization emerging randomly in the region – by defining how often a pixel is selected for possible urbanization. Knowing that Almere is a planned city and the entire municipality is characterized by a very high level of urban planning, this is not surprising – SLEUTH was able to distill this characteristic to a minimal Diffusion parameter. The breed parameter shows significantly more variance, being large for the 10km and 14km extents, while minimal for the larger extents. Looking at the Breed coefficients alone, it is difficult to explain this difference, but interpreting these coefficients together with the Diffusion coefficients paints a clearer picture. The Breed coefficient drives the new spreading center growth, determining the chance that a pixel that has become urbanized during the spontaneous growth phase will become a new spreading center. This dependence on the spontaneous growth phase – and thus on the Diffusion parameter as well – means that if Diffusion is minimal, the value of the Breed coefficient is irrelevant, as there is no spontaneous growth to base the new spreading centers on. Due to path dependency, this can result in largely different coefficients as it did for the current modeling exercise, but this does not affect the outcome either regarding either the spatial fit achieved or the shape of the predicted urbanization pattern.

Slope produced the second-highest variance – similarly to Breed, though this is expectedly caused by path dependence that is a result of the value's irrelevance. Slope signifies the pressure to build upon steeper slopes as well – in other words, Slope determines the effect of steepness on the chance of being urbanized. As a result of being reclaimed from the ocean, Almere is almost completely flat, which is represented in the Slope input layer. This means that whatever the Slope coefficient is, it will have no meaningful effect on the outcome. Road Gravity remains relatively consistent – second-lowest variance after Diffusion – with low-mid range values for extents. Road Gravity defines the effect of Road Influenced growth – the maximum search distance from a road to attract urbanization. Due to the relative weakness of other growth drivers, Edge growth – characterized by the Spread coefficient – remains the main driver of growth for all extents.

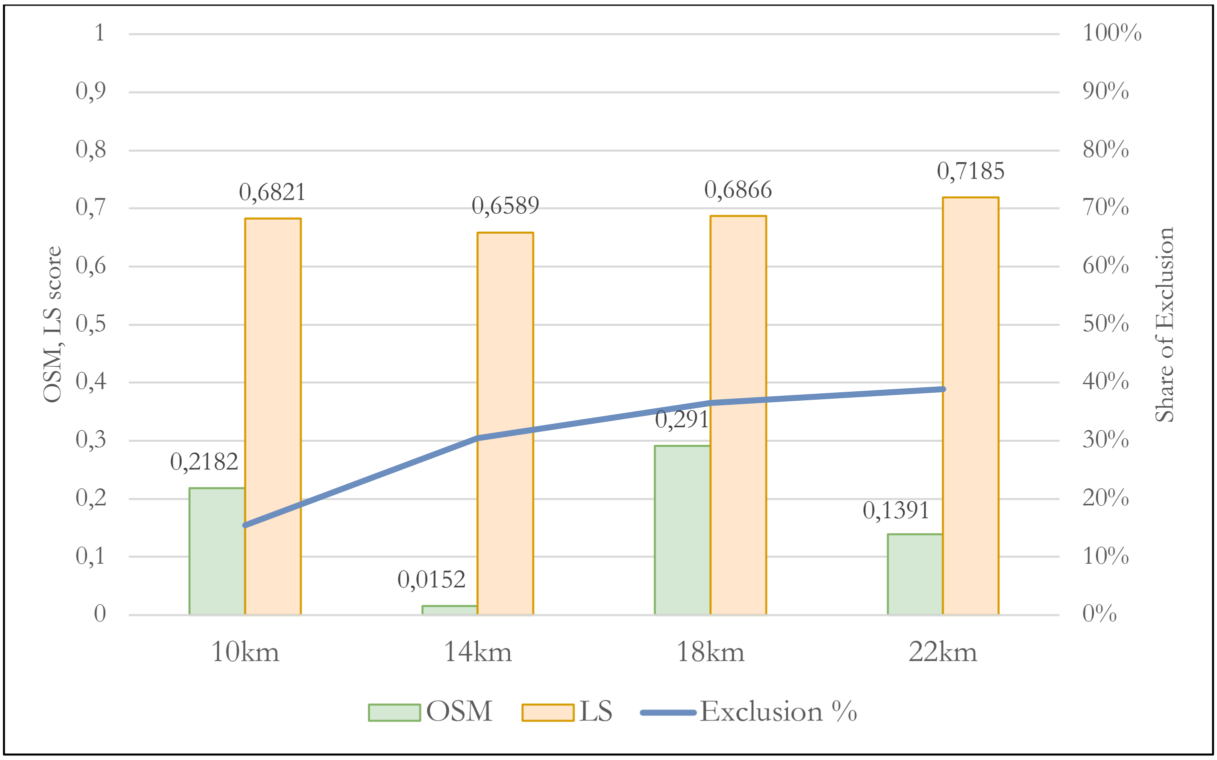
It is worth noting that such interdependence between the coefficients is rarely recognized when assessing the urbanization trends of the future. Partially inspired by a study that found very low Diffusion (2) and high Breed for the Netherlands as a whole, Clarke et al. argue that that the low Diffusion and high Breed coefficients together give a good confirmation of the high level of Dutch regulation for urban planning – new development is not allowed to happen often, but when it does, it is likely to affect a larger area, in a planned manner (Tack, as cited in Clarke et al., 2007, p. 7). This interpretation is in line with the 10km and 14km results for Almere but does not comply with the 18km and 22km results, which achieved low Breed as well. After looking at the prediction results comparing Low Diffusion - Low Breed vs. Low Diffusion – High Breed results, it seems that the high variance of Breed across spatial extents is the result of path dependence, caused by its irrelevance due to low Diffusion. A relatively small – but not minimal – Diffusion and High Breed might be indicative of a high level of spatial planning, but if the Diffusion is minimal, there is nothing for the new spreading center growth to pick up on. These results suggest that one should be careful when inferring insights regarding the 'DNA of a region' after looking at a few coefficients in isolation. These results should always be interpreted in a wider context, taking the finer characteristics of SLEUTH into account as well.

Question 2 – differences of overall spatial fit

Given the different coefficients, it is no wonder that the spatial extents differ in their overall spatial fit as well. Regarding OSM – the metric used for coefficient selection – no overarching pattern can be found, as the best result was achieved by the 18km, and the worst by the 14km extent. This does not reflect any topographical features like the share of Exclusion - which is the largest for the 22km extent, due to the large sea area being present. On the contrary, it might be argued that the 14km extent renders the most 'natural' interpretation of the region around Almere, given that the whole western tip of the Flevopolder island where Almere is located is covered, but the towns south of Gooimeer (the bordering lake separating Flevopolder from the mainland) are still excluded. The Lee-Sallee values hint at a weak trend – larger extents correlate with a slightly better fit – but 14km is still an outlier, and the overall difference in the results is minimal.

Figure 16

OSM and LS scores per extent, compared to the share of Exclusion

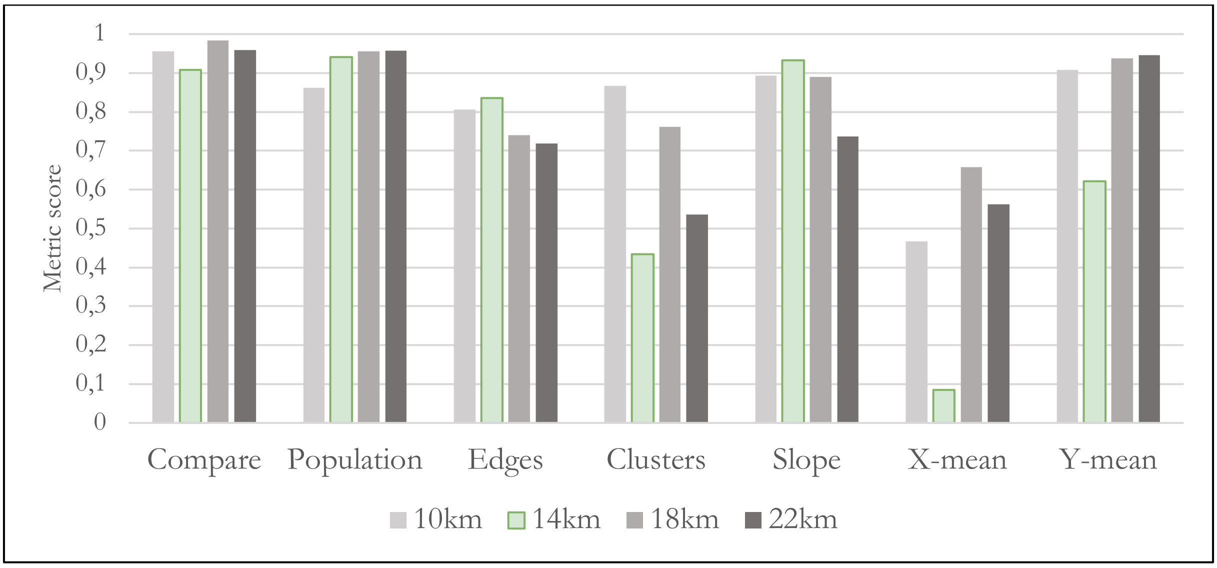


Note. Both OSM and LS scores can range between 0 and 1, with 1 representing perfect spatial fit.

It might be tempting to build conjectures to explain the low spatial fit of 14km, but this question highlights the black box-like aspect of SLEUTH – it is difficult to exactly pinpoint, what causes a divergence in results. Nevertheless, the statistical outputs of the model represent a limited toolset to investigate this issue further. OSM is a composite metric, and by looking at the individual components it can be possible to identify the area where 14km underperformed. OSM is a product of seven metrics. The Compare and Population statistics – similarly to Lee-Sallee – measure the ratio between the number of predicted vs. actually urbanized cells. Edges and Clusters are shape statistics, representing least squares regression scores for the predicted vs. actual edge count and the number of urban clusters. Slope indicates how well the model predicts the slope of urbanized cells compared to the actual ones on average, but given the almost complete flatness of the study area, this metric is not relevant for this current investigation. Lastly, the location statistics X-Mean and Y-Mean are least squares regression scores, representing how well could the model predict the horizontal and vertical locations of the urbanized cells on average. ((Dietzel & Clarke, 2007).

Figure 17

Comparison of the OSM component metrics per extent

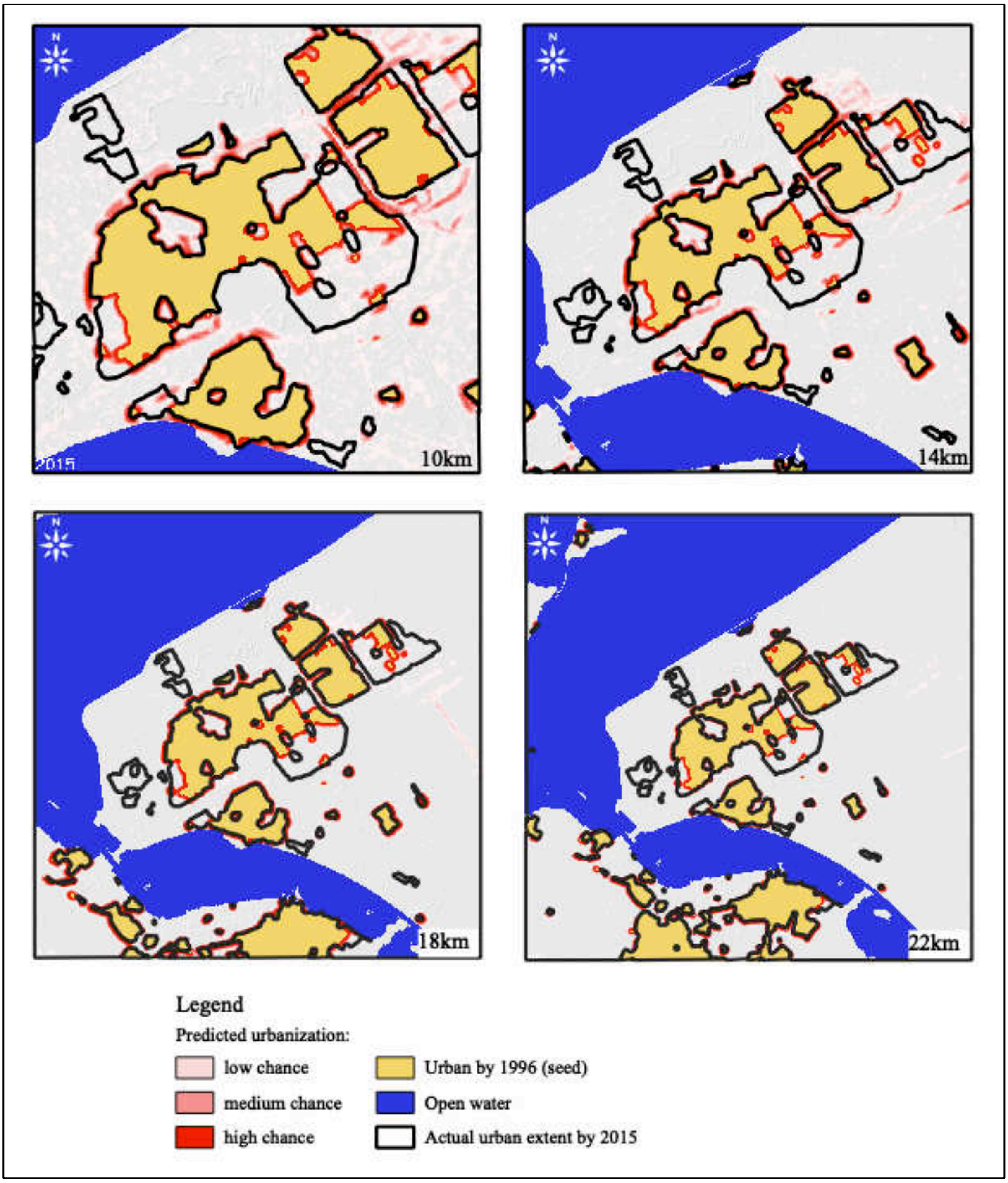


Note. All scores can range between 0 and 1, with 1 representing perfect spatial fit.

The results for Compare, Population, and Edges for 14km are comparable to the other extents. The Clusters metric raises a red flag, indicating that this extent underperformed the others in predicting the location of new urban clusters. Knowing this, it is realistic to expect that the location metrics will be off as well – this is confirmed by the low X-mean and Y-mean scores. Even though both metrics underperform, it is interesting to see that according to the results, 14km especially struggled with predicting the location of urbanization along the East-West axis. This can easily be confirmed by comparing the results visually: for 18km the predicted urbanization is dominantly around the edges, while for 14km (and 10km as well) we can observe a predicted clustering of urban growth along the roads northeast of Almere Buiten – at the northeastern corner of the map. This eastern expansion, however, does not exist, thus decreasing the X-Mean considerably. At this point the limits of investigation to this matter with the tools provided with SLEUTH are reached – it can be deduced *what* caused the lower spatial fit, but there is no indication into *why* was this specific pattern predicted. It is worth noting though that the overall spatial fit for each prediction is relatively low. The aforementioned nonexistent predicted urban extension partially explains the difference in results per extent but is alone not responsible for the overall low spatial fit. Confirming the earlier conjecture, the model struggles to follow the patterns of planned urbanization and misses newly urbanized urban centers such as the ones north and west of Almere Stad.

Figure 18

Comparison of predictions for 2015, compared against actual urban extent

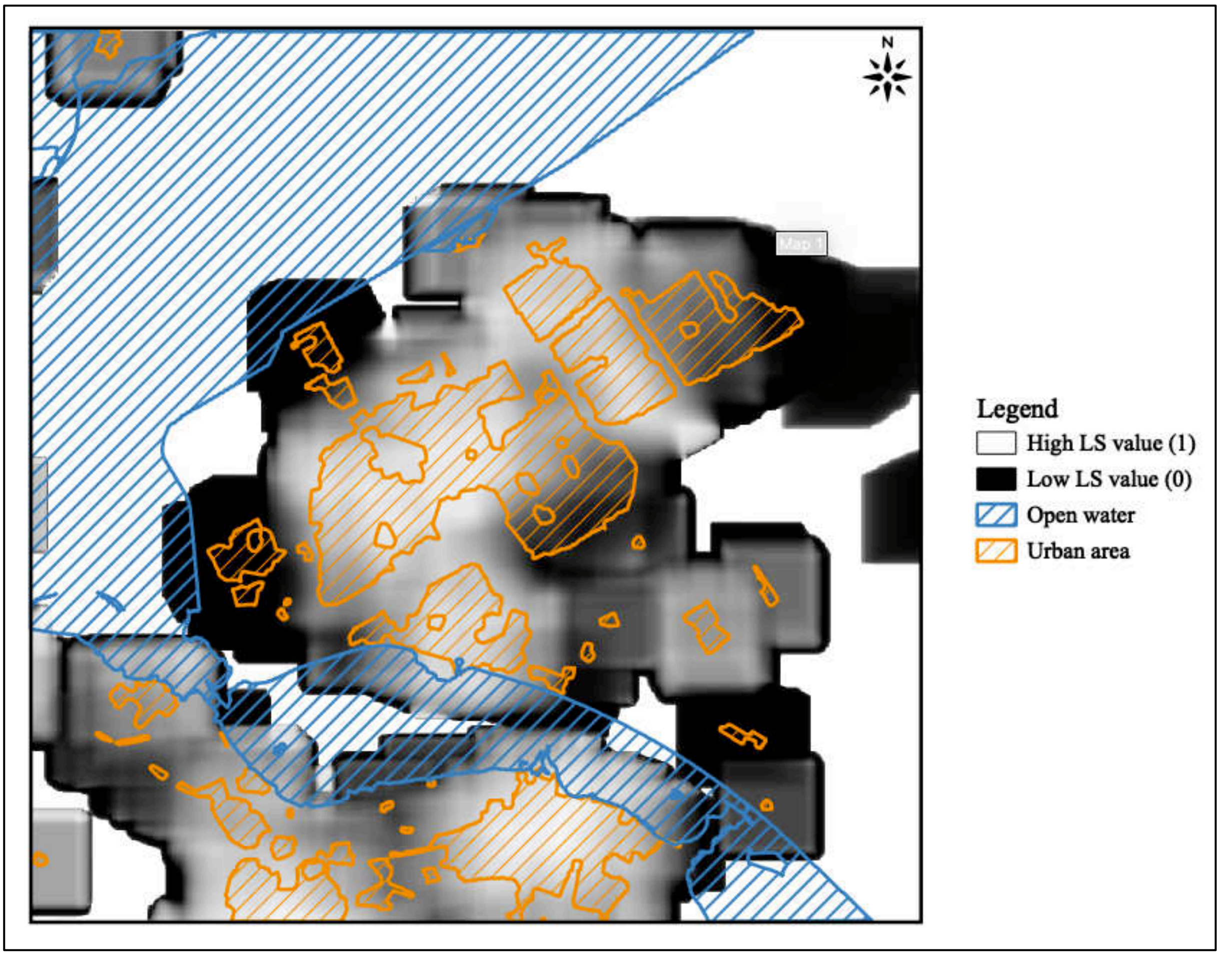


Contrary to OSM, the Lee-Sallee metrics exhibit at least a weak pattern: larger extents seem to correlate with higher scores. The relative simplicity of the metric makes it possible to analyze local trends – at which regions do the metric performs better or worse. Lee-Sallee can be calculated for each pixel based on a given neighborhood radius, thus creating of a map for the 'topography of spatial fit'. The map below indicates this, where the local Lee-Sallee values have been calculated for

22km with 2 km-wide square neighborhoods. White represents the areas with the best local spatial fit. The 1 km-wide band around the edge where the pixels do not have the whole neighborhood available was removed to avoid distortion.

Figure 19

'Topography of spatial fit': local LS values within a 2km neighborhood



The map confirms that the best spatial fit unsurprisingly was achieved along the edges, where the open sea represented on the Exclusion layer guaranteed a high spatial fit. Conversely, the worst spatial fit was achieved around new settlements which appeared after 1996 but the model could not predict them (like Almere Poort, which is the latest addition to the municipality) and around the eastern edge of Almere Buiten, where the model predicted minor urbanization, but it did not happen. The increase in spatial fit per Lee-Sallee correlates with the share of open water – exclusion zones: larger extents represent a larger share of open water as well. This results in three distinct bands for spatial fit:

- A white core area, where the LS values are relatively high. This is due to the high share of areas already urbanized by 1996.
- A black inner ring, where the new urbanization happened, which the model often failed to pick up.
- A white outer ring representing the open water, which is excluded from development.

It must be noted however that the increment in spatial fit is relatively minor, there are only 4 extents to compare and 14km is still an outlier - this investigation alone only indicates a correlation, and should not be taken as a robust proof for the causation between larger Exclusion[6] and better spatial fit. Nevertheless, the correlation is in line with the mechanism of the model: larger Exclusion means less room for SLEUTH to make mistakes, as the extent where the model could incorrectly predict growth is limited.

Question 3 - differences regarding the common extent

It is worth noting that even if a larger Exclusion might increase the overall spatial fit, from a practical perspective this is not necessarily a valid improvement. Including a vast area of ocean as an Exclusion around an island will lead to better a Lee-Sallee metric, but the higher coefficient does not translate into more insightful results. What is more interesting from a practical perspective about comparing predictions with different extents: how does the spatial fit of their intersections match up?

To investigate this question, the common 10km central area of each of the four extents were taken and separate Lee-Sallee metrics were calculated. The four extents produced different predictions even regarding their common intersect, but the distorting effect of different Exclusion sizes is controlled for, as all extents have the same Exclusion for the common 10km central area. Even though the differences are relatively small, the results do exhibit a trend, namely that the larger extents produce better spatial fit. It was previously established that the larger extents produced better overall spatial fit – at least, measured by the Lee-Sallee metric.

[6] See Figure 16 for Exclusion shares per extent.

Table 5

Lee-Sallee metrics, compared by the common 10km extents

10km	0,6267
14km	0,6305
18km	0,6311
22km	0,6360

The topography of spatial fit investigation confirmed that for the current predictions the main accuracy gains were around the edges, representing a larger share of Exclusion as well. Nevertheless, according to the results from the comparison of intersects, larger extents correlate with better spatial fit at the center as well. It is difficult to objectively gauge the correct spatial extent for a given investigation, as often the underlying terms – where do the limits of a city, agglomeration, or region lie – are not clearly defined. But the current results hint at the idea that it is more prudent to err on the 'bigger extent' side: urbanization of the peripheral regions might not be at the focus of a given study, but they might nevertheless influence the development of the core area under focus as well.

CONCLUSION

Summary of findings

This study set out to explore the effect of changing spatial extent on the model outcome, as analyzed by three questions: the effect on coefficients, on overall spatial fit, and the effect on spatial fit measured by the common extent. During the investigation, the whole modeling exercise was executed from data preparation until prediction. As a result of this, insights were gained not only relating to the specified research questions and the defined subquestions but relating to other parts of the modeling process as well – sometimes reinforcing, sometimes contradicting the standing literature. Therefore, the discussion of the conclusions addresses 'core observations' relating to the defined research question and 'stray observations' separately.

Core observations

The first subquestion investigated whether the extension of the study area has any effect on the coefficients. The a priori assumption might be that different inputs will produce different results. There has been no previous study explicitly on this question - this gap in the literature was one of the main inspirations to conduct this study. Nevertheless, researchers have been reporting and comparing coefficients calibrated on different regions, leading to the idea of regional DNA (Silva & Clarke, 2002; Silva, 2004). Even though 'region' lacks a clear definition, the idea that changing the study area marginally can lead to significantly different coefficients puts the theory of regional DNA to a test. This questions the reproducibility - and therefore the comparability of existing studies as well. Nevertheless, a more careful investigation of the coefficient changes revealed patterns relating to underlying data layers. This reinforces the idea that coefficients can and do tell a story - but they always must be interpreted in a wider context, paying attention to the underlying characteristics of SLEUTH as well.

The comparison of spatial fits showed the limitations of the model clearly. The spatial fit measured by LS showed a weak positive correlation to map size, but further investigations revealed that these gains were due to the larger share of Exclusion around the edges, which is more of a technical consequence of the model setup and not a practical insight per se. On the other hand, the results measured by OSM varied significantly without any clear pattern, while 14km underperformed significantly. A deeper look into the OSM component metrics revealed that this divergence is likely due to an incorrectly predicted new urbanization center along the northeastern edge of Almere, which does not exist. These results highlighted the black box-like aspects of SLEUTH: even though careful investigations can reveal what causes the divergence in results (as measured by spatial fit), it provides no insight into why these specific patterns were predicted.

Lastly, the comparison of common spatial extents provided intriguing results. The spatial fit measured by LS displayed a clear, although small positive correlation with map size. This result is especially important as in this scenario the share of Exclusion (the main factor that produced the differences for the previous subquestion) is controlled for. This conclusion highlights the importance of periurban areas for modeling and hints at the idea that even if the focus area of interest is smaller, it makes sense to carry out the modeling on a larger extent.

Stray observations

Non-robustness of GA. The most surprising insight learned while using the model related to the non-robust nature of Genetic Algorithms. Being one of the latest main advancements regarding SLEUTH and showing promising early results, the original research plan relied on GA as the main calibration method. Indeed, GA delivered on increased computational efficiency[7] while achieving

[7] Execution times for both GA and brute-force runs are included in the Empirical Portfolio.

similar results to the brute-force method regarding spatial fit, the outcomes regarding the coefficients were alarmingly non-robust. Running the calibration multiple times with the same model parameters resulted in significantly different coefficients, even though their spatial fit remained comparable. As this study focused on comparative analysis, this represented a critical issue and the fallback brute-force method was used for the actual calibration of the model. It is worth noting that in case the aim of the modeling exercise is the prediction itself – with a good spatial fit – and the modeler is not interested in the underlying coefficients per se, this might be less of an issue. Nevertheless, this highlights a gap in our understanding regarding GA and its application. If GA is to live up to its promises and see proliferation as the new de-facto calibration method, further analysis on this matter is highly required.

Unsuitability for high level of planning. The overall spatial fit achieved at the end of the calibration, measured by either metrics, is relatively low for all extents. Given that this study focused on the comparison of outcomes and not on optimizing results for a better spatial fit, this is not a critical issue. Nevertheless, it must be acknowledged that the LS and OSM values are lower than the average in SLEUTH literature. A deeper investigation revealed that this is primarily due to the model's inability to follow urbanization trends when they are centrally planned, instead of happening organically. This is not entirely in line with the existing corpus: Silva and Clarke (2002) emphasize that the model is universally portable and highlight its application to European cities, while Al-shalabi et al. (2013) explicitly confirm that SLEUTH can be applied to highly planned and organized regions as well. However, in a recent study using a different model Ferdinand (2020) finds that spatial fit for European regions tends to significantly underperform other continents, as forecasting models become less accurate for highly developed urban areas. The findings of this study are more in line with Ferdinand's, as the model could not pick up the centrally planned urbanization patterns characterizing the growth of Almere.

Limits of the regional DNA. The idea that the coefficients of the model can tell a story about the underlying urbanization trends of the modeled region (Silva & Clarke, 2002; Silva 2004) was the focus of investigations for the research project that was the forerunner of this thesis. The current investigations however showed the limits of this approach: it is problematic to define what constitutes the 'DNA of region' if a small change to the extent results in significantly different coefficients. At the heart of this problem lies a still unsolved question, namely that we miss a commonly accepted definition for 'region', and how the study area should be delineated for modeling. Nevertheless, a more detailed analysis of the results revealed that the coefficients did not change randomly from extent the extent: their divergence can be explained by looking at the underlying data and the mechanisms of the model. But this highlights that alone it is usually not sufficient to look at a single set of coefficients to infer the underlying trends of urbanization. These coefficients should be analyzed more prudently, always in context, with considering the finer attributes of the model as well.

Prevalence of discretionary choices. Clarke (2004) mentions that systems need to complex, but not too complex. This applies to the process of spatial modeling as well, and Jafarnezhad et al. (2015) suggest that a significant step towards the desirable balance is to decrease the number of human decisions needed during the calibration process. Discretionary choices not only dominate the calibration process though, but they are also unavoidable during the definition phase (selection of extent, resolution, timescale; land classification; etc.) as well. On one hand, this undermines the comparability of results between studies. But on the other, the lack of a common research protocol means that the current method of modeling is more susceptible to 'forced results' as well. As an example, this study showed that the model is very sensitive to minor changes in the Exclusion layer, but there is no common understanding or guideline about how the Exclusion layer should be set up. This makes it very easy to engineer the Exclusion specifically to a targeted outcome, which nevertheless remains very hard to scrutinize. But one does not need to act in bad faith, as significantly different outcomes can be the result of seemingly trivial different human decisions as well. To combat this, as Jafarnezhad et al. (2015) suggested, it is necessary to gradually replace subjective human decisions with objective processes. But until then, a high level of honesty during the modeling exercise has to be presumed.

Limitations

Scope. It is important to highlight that the study focused only on one area, within a limited period, comparing only four different extents. This is also the main reason no statistical modeling was applied to examine the correlation between extent size and coefficients / spatial fit: the limited scope of the research yielded four data points (for four extents), which is not a sufficient base for robust comparisons. This of course limits the interpretation of the results as well. The current results are intended to serve as an indication and as a blueprint for further investigations. But to reach more cogent conclusions, further studies on a larger scope are needed.

Comparability. As mentioned previously, due to the discretionary choices the model requires the results (either the coefficients or the spatial fit) are hardly comparable between studies. This makes the interpretation of results even harder, as there is no benchmark to compare against. Take for instance the change of spatial fit metrics per extent (Figure 16): due to the lack of valid comparisons, it is difficult to gauge whether the results in absolute terms are good or bad, or whether the difference between the extents is significant. This limitation applies not only to this study but to the wider body of SLEUTH literature as well, as currently we lack a common protocol that facilitates the creation of results that are conducive to comparison.

Protocol. As argued previously, the limitations on comparability stem from the lack of an existing common research protocol to follow. This deficiency, however, affects the internal validity of the experiment too: as the model relies on human decisions, it is difficult to objectively gauge whether a decision made during the modeling exercise (whether it relates to the definition, calibration, or prediction phase) was appropriate or not. It must be noted though that this affects not only the modeler but the reader or reviewer as well, as it is difficult to evaluate whether a given decision was valid and in integrity with the study. This study put extra emphasis on explaining the choices made during the definition and calibration phase (including the data preparation process) to provide greater transparency into these details.

Further studies

This paper purposefully focused only a narrowly defined question that so far was not in the focus of SLEUTH-related studies: the model's sensitivity to spatial extent. Nevertheless, this study only aims to kickstart the discussions related to this question. The results have shown that this often-overlooked aspect of the modeling exercise does affect the outcome significantly.

Given the limited resources available, the scope of investigations was constrained as well. To better understand the effect of scope on the model outcome, more research into this question is necessary. One way to do this is to extend the focus of the study.

- *Other study areas.* Almere is special in the sense that it is a planned city, and the results confirmed that SLEUTH struggles with modeling regions with a high level of urban planning. Modeling other regions (especially where central planning is less pronounced) can provide further insights.
- *More extents.* This paper only explored four different extents, which does not lend itself to robust spatial analysis. Comparisons with more map sizes can help understand the finer trends related to the change of extent. An applicable extension of the current approach

would be to define scenario-based extents ('core city', 'agglomeration', 'basin', etc.), instead of the current rule-based model ('consecutive extents increase with 2km') for comparison. This might provide insights that could be readily applicable for other studies as well.

- *Different spatio-temporal setups.* The study examined only an 18km area, on a 25m resolution, within a two-decade period, with control data available for roughly every third year. To achieve more generally applicable conclusions, investigations on a larger scope are needed. As an example, it should be highlighted that even the largest extent (18km) modeled is relatively small compared to other studies utilizing SLEUTH (Dietzel & Clarke, 2004). Further studies utilizing significantly bigger map sizes for comparison might yield different insights.

Among the four Spatio-temporal factors mentions – map extent, map resolution, period, time step – this study focused on map extent as this question is so far the least discussed in the SLEUTH literature. Nevertheless, this does not mean that the questions regarding the remaining three factors are solved, or that we have a general understanding of their sensitivity related to the model outcome. To make appropriate modeling decisions regarding these factors, we need a deeper understanding of how they work – and how they affect each other – warranting further research into these questions.

Last but not least, the author would like to pinpoint the issue that he believes to be the biggest shortcoming plaguing the SLEUTH corpus as of 2021: the lack of a common research protocol, leading to arbitrary human decisions during the modeling, which results in weak reproducibility and the inability to compare results across studies. More stringent documentation of the decisions and steps made during the modeling exercise can partially alleviate this problem. However, until the research community does not have a common protocol – or to paraphrase Goldstein (2004), until we manage to replace the brawn with the brain – using distant methods, we are all stuck in our separate ivory towers.

REFERENCES

Abd-Allah, M. A., & Mohamed, M. (2007). Modelling urban dynamics using geographic information systems, remote sensing and urban growth models. Doctoral Dissertation. *Faculty of Engineering at Cairo University, Giza.*

Al-shalabi, M., Pradhan, B., Billa, L., Mansor, S., & Althuwaynee, O. F. (2013). Manifestation of remote sensing data in modeling urban sprawl using the SLEUTH model and brute force calibration: a case study of Sana'a city, Yemen. *Journal of the Indian Society of Remote Sensing*, *41*(2), 405-416.

Bihamta, N., Soffianian, A., Fakheran, S., & Gholamalifard, M. (2015). Using the SLEUTH urban growth model to simulate future urban expansion of the Isfahan metropolitan area, Iran. *Journal of the Indian Society of Remote Sensing*, *43*(2), 407-414.

Candau, J. (2000). Calibrating a cellular automaton model of urban growth in a timely manner. In *Proceedings of the 4th international conference on integrating geographic information systems and environmental modeling: problems, prospects, and needs for research* (pp. 2-8). University of Colorado.

Candau, J. (2002). Temporal calibration sensitivity of the SLEUTH urban growth model. Doctoral Dissertation. *University of California, Santa Barbara.*

Chaudhuri, G., & Clarke, K. (2013). The SLEUTH land use change model: A review. *Environmental Resources Research*, *1*(1), 88-105.

Clarke-Lauer, M. D., & Clarke, K. C. (2011). Evolving simulation modeling: Calibrating SLEUTH using a genetic algorithm. In *Proceedings of the 11th International Conference on GeoComputation, London, UK* (Vol. 2022, pp. 20-22).

Clarke, K. C. (2004). Toward GeoComputational Honesty in Urban Modeling. *GeoDynamics*, 215.

Clarke, K. C. (2008). A decade of cellular urban modeling with SLEUTH: Unresolved issues and problems. *Ch*, *3*, 47-60.

Clarke, K. C. (2018). Land use change modeling with SLEUTH: improving calibration with a genetic algorithm. In *Geomatic approaches for modeling land change scenarios* (pp. 139-161). Springer, Cham.

Clarke, K. C., Gazulis, N., Dietzel, C., & Goldstein, N. C. (2007). A decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model. *Classics in IJGIS: twenty years of the international journal of geographical information science and systems*, 413-427.

Clarke, K. C., 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*(2), 247-261.

Dadashpoor, H., & Nateghi, M. (2017). Simulating spatial pattern of urban growth using GIS-based SLEUTH model: a case study of eastern corridor of Tehran metropolitan region, Iran. *Environment, Development and Sustainability*, *19*(2), 527-547.

Dezhkam, S., Amiri, B. J., Darvishsefat, A. A., & Sakieh, Y. (2014). Simulating the urban growth dimensions and scenario prediction through sleuth model: a case study of Rasht County, Guilan, Iran. *GeoJournal*, *79*(5), 591-604.

Dietzel, C., & Clarke, K. C. (2004). Spatial differences in multi-resolution urban automata modeling. *Transactions in GIS*, *8*(4), 479-492.

Dietzel, C., & Clarke, K. C. (2007). Toward optimal calibration of the SLEUTH land use change model. *Transactions in GIS*, *11*(1), 29-45.

Duivesteijn, A. (2007). Schaalsprong Almere 2030. Randstad Urgent. http://www.adriduivesteijn.nl/wp-content/uploads/Contract-Randstad-Urgent.pdf

Ferdinand, P. (2020). Forecasting patterns of urban expansion. Master's Thesis. Retrieved from https://spinlab.vu.nl/wp-content/uploads/2020/09/PendulaFerdinand_Forecasting-patterns-of-urban-expansion.pdf

Gandhi, S., & Suresh, V. (2012). Prediction of Urban Sprawl in Hyderabad City using Spatial Model, Remote Sensing and GIS Techniques. *International journal of scientific research, 1*, 97-100.

Gemeente Almere. (2021). Number of inhabitants of Almere. https://www.almere.nl/over-almere/feiten-en-cijfers/bevolkingsprognose-1

Goldstein, N. C. (2004). Brains versus brawn-comparative strategies for the calibration of a cellular automata-based urban growth model. In *GeoDynamics* (pp. 249-272). CRC Press.

Hua, L., Tang, L., Cui, S., & Yin, K. (2014). Simulating urban growth using the SLEUTH model in a coastal peri-urban district in China. *Sustainability*, *6*(6), 3899-3914.

Hua, L., Tang, L., Cui, S., & Yin, K. (2014). Simulating urban growth using the SLEUTH model in a coastal peri-urban district in China. *Sustainability*, *6*(6), 3899-3914.

Jafarnezhad, J., Salmanmahiny, A., & Sakieh, Y. (2016). Subjectivity versus objectivity: comparative study between brute force method and genetic algorithm for calibrating the SLEUTH urban growth model. *Journal of Urban Planning and Development*, *142*(3), 05015015.

Jantz, C. A., & Goetz, S. J. (2005). Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science*, *19*(2), 217-241.

Jantz, C. A., Goetz, S. J., & Shelley, M. K. (2004). Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the Baltimore-Washington metropolitan area. *Environment and Planning B: Planning and Design*, *31*(2), 251-271.

Jantz, C. A., Goetz, S. J., 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), 1-16.

KantaKumar, L. N., Sawant, N. G., & Kumar, S. (2011). Forecasting urban growth based on GIS, RS and SLEUTH model in Pune metropolitan area. *International journal of geomatics and geosciences*, *2*(2), 568.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464-1480.

Lee, D. R., & Sallee, G. T. (1970). A method of measuring shape. *Geographical Review*, 555-563.

Liu, X., Sun, R., Yang, Q., Su, G., & Qi, W. (2012). Simulating urban expansion using an improved SLEUTH model. *Journal of Applied Remote Sensing*, *6*(1), 061709.

Liu, Y., & Liu, X. (2009). Applying SLEUTH for simulating urban expansion of Hangzhou. In *Second International Conference on Earth Observation for Global Changes* (Vol. 7471, p. 747104). International Society for Optics and Photonics.

Liu, Y., Li, L., Chen, L., Cheng, L., Zhou, X., Cui, Y., ... & Liu, W. (2019). Urban growth simulation in different scenarios using the SLEUTH model: A case study of Hefei, East China. *Plos one*, *14*(11), e0224998.

Mahiny, A. S., & Gholamalifard, M. (2011). Linking SLEUTH urban growth modeling to multi criteria evaluation for a dynamic allocation of sites to landfill. In *International Conference on Computational Science and Its Applications* (pp. 32-43). Springer, Berlin, Heidelberg.

Obeysekera, J., & Rutchey, K. (1997). Selection of scale for Everglades landscape models. *Landscape Ecology*, *12*(1), 7-18.

Rafiee, R., Mahiny, A. S., Khorasani, N., Darvishsefat, A. A., & Danekar, A. (2009). Simulating urban growth in Mashad City, Iran through the SLEUTH model (UGM). *Cities*, *26*(1), 19-26.

Sakieh, Y., Amiri, B. J., Danekar, A., Feghhi, J., & Dezhkam, S. (2015). Simulating urban expansion and scenario prediction using a cellular automata urban growth model, SLEUTH, through a case study of Karaj City, Iran. *Journal of Housing and the Built Environment*, *30*(4), 591-611.

Sakieh, Y., Salmanmahiny, A., Jafarnezhad, J., Mehri, A., Kamyab, H., & Galdavi, S. (2015). Evaluating the strategy of decentralized urban land-use planning in a developing region. *Land use policy*, *48*, 534-551.

Sangawongse, S., Sun, C. H., & Tsai, B. W. (2005, December). Urban growth and land cover change in Chiang Mai and Taipei: results from the SLEUTH model. In *Proceedings of MODSIM* (pp. 2622-2628).

Saxena, A., & Jat, M. K. (2019). Capturing heterogeneous urban growth using SLEUTH model. *Remote Sensing Applications: Society and Environment*, *13*, 426-434.

Silva, E. A. (2004). The DNA of our regions: artificial intelligence in regional planning. *Futures*, *36*(10), 1077-1094.

Silva, E. A., & Clarke, K. C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, environment and urban systems*, *26*(6), 525-552.

Silva, E. A., & Clarke, K. C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, environment and urban systems*, *26*(6), 525-552.

Tan, L., Li, M. C., Liu, Y. X., Wei, W., & Hu, W. (2009). Study of urban expansion simulation on the condition of ecological environment protection:—A case study in Dianchi Basin in Kunming. In *2009 Joint Urban Remote Sensing Event* (pp. 1-6). IEEE.

Venema, D. J. (2016). Modelling Land Cover Change Using Geographic Information Systems, Remote Sensing, an Urbanization-suitability Layer and SLEUTH-3r for Groningen, the Netherlands. Master's Thesis. Retrieved from https://research.tue.nl/files/60889093/Venema_0868690.pdf

Watkiss, B. M. (2008). The SLEUTH urban growth model as forecasting and decision-making tool. Doctoral Dissertation. *Stellenbosch University, Stellenbosch.*

Xi, F., He, H. S., Hu, Y., Wu, X., Bu, R., Chang, Y., ... & Yu, J. (2009). Simulate urban growth based on RS, GIS, and SLEUTH model in Shenyang-Fushun metropolitan area northeastern China. In *2009 Joint Urban Remote Sensing Event* (pp. 1-10). IEEE.

Xibao, X., Feng, Z., & Jianming, Z. (2006). Modelling the impacts of different policy scenarios on urban growth in Lanzhou with remote sensing and Cellular Automata. In *2006 IEEE International Symposium on Geoscience and Remote Sensing* (pp. 1435-1438). IEEE

Yang, X., & Lo, C. P. (2003). Modelling urban growth and landscape changes in the Atlanta metropolitan area. *International journal of geographical information science*, *17*(5), 463-488.

Yin, C., Yu, D., Zhang, H., You, S., & Chen, G. (2008). Simulation of urban growth using a cellular automata-based model in a developing nation's region. In *Geoinformatics 2008 and joint conference on GIS and built environment: Geo-simulation and virtual GIS environments* (Vol. 7143, p. 714314). International Society for Optics and Photonics.

APPENDIX

Table 6

Urban Extent Classification Table

BBG Code	Label (Dutch)	Label (English)	Urban
10	spoorwegen	railways	0
11	hoofdweg	highway	0
20	woongebied	residential area	1
21	detailhandel en horeca	retail and catering	1
22	openbare voorzieningen	public facilities	1
23	sociaal-culturele voorzieningen	sociocultural facilities	1
24	bedrijfsterreinen	business premises	1
30	stortplaatsen	landfills	0
32	begraafplaats	cemetery	1
34	bouwterrein	building site	0
35	semi verhard overig terrein	semi-paved other terrain	0
40	parken en plantsoenen	parks and public gardens	1
41	sportterreinen	sports grounds	1
42	volkstuinen	allotments	1
43	dagrecreatieve terreinen	day recreation areas	1
44	verblijfsrecreatie	residence recreation	1
50	glastuinbouw	greenhouse horticulture	0
51	overige agrarisch gebruik	other agricultural use	0
60	bos	forest	0
61	droog natuurlijk terrein	dry natural terrain	0
62	nat natuurlijk terrein	wet natural terrain	0
70	IJsselmeer/Markermeer	IJsselmeer / Markermeer	0
73	Randmeer	Randmeer	0
75	water met een recreatieve functie	water with a recreational function	0
78	overig binnenwater	other inland waterways	0

Note. Only classes that are present within the 24km study area are listed. The description of land cover classes by CBS is included in the Empirical Portfolio (in Dutch).

Table 7

Exclusion Classification Table

BBG Code	Label (Dutch)	Label (English)	Exclusion
10	spoorwegen	railways	0
11	hoofdweg	highway	0
20	woongebied	residential area	0
21	detailhandel en horeca	retail and catering	0
22	openbare voorzieningen	public facilities	0
23	sociaal-culturele voorzieningen	sociocultural facilities	0
24	bedrijfsterreinen	business premises	0
30	stortplaatsen	landfills	0
32	begraafplaats	cemetery	0
34	bouwterrein	building site	0
35	semi verhard overig terrein	semi-paved other terrain	0
40	parken en plantsoenen	parks and public gardens	0
41	sportterreinen	sports grounds	0
42	volkstuinen	allotments	0
43	dagrecreatieve terreinen	day recreation areas	0
44	verblijfsrecreatie	residence recreation	0
50	glastuinbouw	greenhouse horticulture	0
51	overige agrarisch gebruik	other agricultural use	0
60	bos	forest	0
61	droog natuurlijk terrein	dry natural terrain	0
62	nat natuurlijk terrein	wet natural terrain	0
70	IJsselmeer/Markermeer	IJsselmeer / Markermeer	100
73	Randmeer	Randmeer	100
75	water met een recreatieve functie	water with a recreational function	100
78	overig binnenwater	other inland waterways	100

Note. Exclusion values range from 0 (non-excluded) to 100 (excluded). Only classes that are present within the 24km study area are listed. The description of land cover classes by CBS is included in the Empirical Portfolio (in Dutch).

Table 8

Water Overlay Classification Table

BBG Code	Label (Dutch)	Label (English)	Water Overlay
70	IJsselmeer/Markermeer	IJsselmeer / Markermeer	1
73	Randmeer	Randmeer	1
75	water met een recreatieve functie	water with a recreational function	0
78	overig binnenwater	other inland waterways	0

Note. Only classes that are part of the Exclusion layer (

Table 7) are included. The description of land cover classes by CBS is included in the Empirical Portfolio (in Dutch).

Table 9

Road Weight Table

Type	Weight
motorway	80
primary	40
secondary	20
tertiary	10
bridleway	0
cycleway	0
footway	0
living_street	0
motorway_link	0
path	0
pedestrian	0
primary_link	0
residential	0
service	0
steps	0
track	0
trunk	0
unclassified	0

Note. Road weight values range from 0 (no road) to 100 (highest level road). Only classes that are present within the 24km study area are listed.