

Improving the modelling of land-use changes due to transport system characteristics in Metronamica

Joël Meijers

16-8-2012

Research Institute for Knowledge Systems

&

University of Twente

Civil Engineering & Management

Improving the modelling of land-use changes due to transport system characteristics in Metronamica

Master thesis report

J.H.J. Meijers

Supervisors

Dr. Ing. K.T. Geurs University of Twente

T. Brands MSc. University of Twente

Ir. H. van Delden RIKS

16-8-2012

Enschede, the Netherlands

Version 9 (final)



UNIVERSITY OF TWENTE.

Preface

This is written in the final hours. Not only the final hours before this report's delivery deadline, but also the final hours of a five month research at RIKS in Maastricht. These five months are characterised by a challenging research, lots of back-and-forth travelling and a bit too much of screen time.

The final hours of my study Civil Engineering at the University of Twente have also begun by finishing this report. The past six years have been about everything that connects science and policy. In this sense, RIKS was a perfect place to do my Master thesis, since the link to policy is so eminent.

I would like to thank everybody who contributed to this report by offering me one or more of the following things: feedback on my research progression, food for thought, a place to stay during my time in Maastricht, patience, one or more piece(s) of 'vlaai' and some good laughs.

I hope that reading this report will prove to be both enjoyable and valuable to you.

Enschede, 22-7-2012

Joël Meijers

Summary

"We now know a lot about the interactions of transport and land-use, but we also know very little" (Levinson, 2011). The relationship between transport and land-use is generally seen as an interactive relationship (Wegener & Fürst, 1999). Many theoretical models try to capture this relationship by introducing intermediate variables such as economic productivity and accessibility (Bruinsma, 1994; Zondag, 2007). In general, accessibility is seen as the major intermediate variable: the transport system determines the accessibility of locations, which in turn determines where land-use developments are going to take place. Accessibility can be defined as 'the extent to which land-use and infrastructure systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s)' (Geurs & Van Wee, 2004).

Most empirical research on land-use transport interaction has focussed on the effects of land-use patterns on transport systems (Wegener & Fürst, 1999). Research that investigates the opposite effect is more sparse (e.g. Levinson, 2008) and the findings contradict each other. This calls for more empirical evidence to gain further insight in the transport effects on land-use developments. This research focuses on the impact of the location and quality of transport elements (railway stations, motorway ramps and roads) on land-use changes in the Netherlands in the period 1989-2000. The goal is to improve the Metronamica LUT modelling of land-use changes due to transport system characteristics by providing empirical evidence and using this evidence to refine and calibrate the accessibility measure.

Metronamica LUT is a land-use development simulation model with a cellular automata basis developed by the Research Institute for Knowledge Systems (RIKS). The land-use simulation is driven by four drivers: neighbourhood dynamics, zoning plans, physical suitability and accessibility. The latter consists of several components. The research focuses on the 'local accessibility' component, which is basically a measure for the distance to the nearest transport element.

The research contains two parts. Firstly, it provides a data analysis on the proximity of land-use changes to transport elements, for which overrepresentation graphs are used. Secondly, it investigates whether and how these findings can be used in the calibration procedure of Metronamica LUT and whether differentiating between transport element quality improves the land-use simulation.

Data analysis

The data analysis investigates whether certain land-use developments are overrepresented near certain transport elements for all land-use type-transport element combinations. Overrepresentation is defined as the difference between the expected density of land-use developments (according to a homogeneous distribution over the study area) and the observed density in the neighbourhood of transport elements. The overrepresentation is plotted to distance in overrepresentation graphs.

It was chosen to work with these graphs, since they can easily be compared with similar graphs for attraction of land-use development to existing land-use types, which are already available (Verburg, de Nijs, et al., 2004). Another advantage is that the overrepresentation graphs can easily be translated into distance decay parameters which are used in the local accessibility component of Metronamica LUT. A third, more practical reason is that creating a tool for handling the necessary calculations took relatively little time.

For railway stations, quality is defined by whether or not the station is attended by intercity trains. This relatively simple classification was chosen to see whether notable differences in overrepresentation would arise. Note that this is useful information, sine more specific data may not always be available. Besides, we are most interested in the trend, i.e. how overrepresentation changes when moving from 'small' to 'big' stations. For this reason, the exact classification is not that important.

For roads, three different traffic characteristics, used for defining quality, were studied separately: flow (veh/h), congestion (flow/capacity ratio) and operating speed (km/h). For each of these characteristics, three quality levels were defined: low, medium and high. The classification of values into these three categories was based on a Jenks natural breaks method, since this method suits an uneven distribution of values. Overrepresentation graphs were created for all three traffic characteristics, in order to see which of these is most suited to explain land-use developments. Ramps are not further classified based on quality, since no data are available.

It was found in the data analysis that not only the location of transport elements, but also the quality of the element makes a difference in the extent to which land-use developments are overrepresented near the element. The overrepresentation of commercial developments near intercity stations is much larger than near local stations. Similarly, commercial developments are seen more around low flow highways than high flow highways. For residential developments, it is the other way around. Although most relations were expected by theory, the overrepresentation near stations (clearly indicated in this research) had not had many back-up from other research. The overrepresentation itself, however, does not necessarily indicate a direct causal relationship between the transport element and the land-use development. Therefore, every outcome of the data analysis must be interpreted carefully.

Modelling

The second part of the research focussed on improving Metronamica LUT's land-use development simulation for the Netherlands. The effects of introducing differentiation to transport element quality in the model were tested. This differentiation is reached by setting different distance decay parameters for different quality levels of the same transport element.

Three simulations were carried out. The first did not include the accessibility driver. The land-use simulation is thus only determined by the other drivers in Metronamica LUT: neighbourhood dynamics, zoning plans and physical suitability. The second simulation included accessibility without different parameter values for different quality levels. The third simulation entailed a slightly adjusted accessibility measure that enables the user to set different parameters values for different quality levels of the same transport element. These three simulations were set up in this particular way to be able to test the effects of the inclusion of the accessibility driver in the model (comparison between simulation I and II) and also the effects of the inclusion of transport element quality differentiation (comparison between simulation II and III).

For railway stations, quality was defined the same as in the data analysis (intercity and local stations). For roads, however, quality is defined as the flow rather than the congestion or speed, because the latter two depend on an extra variable: the capacity of the road. Data on the capacity are not always reliable (for example on small links that represent motorway ramps) and would thus introduce another source of error. This can be seen in the fact that the flow/capacity ratios in the simulation reach extreme high values. Besides, the data analysis shows a stronger link with flow than congestion. The simulated flow is used to dynamically categorise roads in three quality levels. However, since the data analysis showed that overrepresentation drastically decreases for high congestion levels, all roads with a congestion ratio higher than 1 are defined as a separate transport element (without further quality differentiation).

A set of indicators is used to assess the simulations. Three types of indicators should be used (White, Engelen, & Uljee, 1997), namely indicators for location agreement (Kappa variants), indicators for pattern agreement (e.g. cluster size-frequency distribution and overrepresentation graphs) and, perhaps the most powerful indicator, visual inspection. It was found that besides visual inspection of the maps, especially the overrepresentation graphs were a useful indicator to guide the calibration of the local accessibility parameters. This is no surprise, since the overrepresentation graphs are translated into attraction rules in the local accessibility parameters and, the other way around, adjusting the parameters leads to a direct influence of the overrepresentation graphs.

The first simulation showed that some of the overrepresentation along transport elements was already present in a simulation that does not include attraction rules to infrastructure. This is a clear hint that not all overrepresentation that was shown in the data analysis, has a direct causal relationship with the element. On the other hand, it is also not ruled out that the neighbourhood rules model some processes that are caused by the transport element in reality. This confirms once more that causality remains a complex issue in land-use modelling.

After including and calibrating accessibility in the other simulations, the way land-use patterns are simulated generally improved, but the number of cells that are modelled correctly decreased. This can be seen both visually and by examining the numerical indicators mentioned above (of which some evaluate locations and some evaluate patterns). Including a differentiation between transport elements quality, strengthens this effect. The question that arises here is whether the simulation has become better or worse. This remains a subjective assessment, although it must be said that it is not in the nature of Metronamica LUT to exactly model land-use changes, but rather to generate the general patterns by modelling underlying processes. In this sense, including accessibility and differentiation between transport element quality improves the model. At the very least, it enables the modeller to find a balance between fitting to location and to pattern, which is appropriate for the specific model application.

The changes that were caused by adjusting the local accessibility parameters were smaller than expected, for which several reasons can be mentioned. These reasons are also important considerations to keep in mind when analysing the data analysis results. First of all, zoning plans play traditionally a large role in land-use developments of the Netherlands, which was not different during the 90s (for example, large housing development sites were determined in Vinex). Another reason is that the Netherlands has a consistently well-developed infrastructure network over the whole country. The influence of accessibility on location choices is therefore expected to be of a smaller extent than in other areas. For other study areas, it is expected that setting different parameter values for different quality levels will be of greater use and that the location and pattern indicators might not contradict each other. Yet another reason might lie in the used data for the Netherlands: changes in both the scale level (resolution) and the land-use type classification might yield different results.

Contents

1	Research.....	9
1.1	Goal.....	9
1.2	Transport, accessibility and land-use	10
1.3	Research questions.....	10
1.4	Report outline.....	11
2	Theory.....	12
2.1	Land-use transport interaction theories	12
2.2	Empirical findings.....	14
2.3	Empirical findings for the Netherlands.....	16
2.4	Accessibility	17
2.5	Conclusions	18
2.6	Implications for Metronamica LUT	19
3	Metronamica LUT description.....	20
3.1	Introduction to Metronamica LUT	20
3.2	Transport model	21
3.3	Land-use model	21
3.4	Neighbourhood dynamics.....	22
3.5	Accessibility	23
3.6	Metronamica in this research	26
4	Data	27
4.1	Sources	27
4.2	Land-use data	27
4.3	Transport network data	32
4.4	Model application	33
4.5	Data usage throughout this research.....	33
5	Data analysis	35
5.1	Methodology	35
5.2	Setup	40
5.3	Results.....	41
5.4	Discussion.....	54
5.5	Conclusions	55
5.6	Reflection.....	56
6	Local accessibility calibration and validation.....	58
6.1	Simulation set-up	58
6.2	Calibration and validation procedure	60
6.3	Results & discussion.....	62

6.4	Conclusions	64
6.5	Reflection.....	65
7	Conclusion	71
7.1	Research question 1	71
7.2	Research question 2.....	71
7.3	Research question 3	72
7.4	Discussion.....	72
8	References.....	73
	Appendix A: Data description	75
	Appendix B: Road classification intervals	77
	Appendix C: Region maps	78
	Appendix D: Example of data analysis Excel sheet.....	79
	Appendix E: Flowchart of data analysis process	80
	Appendix F: Vinex map.....	81
	Appendix G: Implication of overlapping observation areas	82
	Appendix H: Settings of other parameters in simulations.....	85
	Appendix I: Adjusted local accessibility measure	86
	Appendix J: Probability maps.....	87
	Appendix K: Enlarged maps.....	89

1 Research

The spatial planning support system Metronamica LUT is used to explore land-use developments and the impact of policy on these developments. Many determinants can explain land-use developments to a certain degree (e.g. Iacono & Levinson, 2008; Verburg, Ritsema van Eck, de Nijs, Dijst, & Schot, 2004), one of which is transport. The relationship between land-use and transport is mostly viewed as being interactive, which is famously depicted by the land-use transport feedback cycle in Figure 1.1 (Wegener & Fürst, 1999).

In the feedback cycle, accessibility is seen as the key intermediate variable for explaining the effects of (changes in) the transport system on land-use developments, although numerous other options have been proposed, like effects due to agglomeration and economic productivity (Bruinsma, 1994). To add yet another layer of complexity, it is crucial to realise that 'accessibility' is a broad term that can be defined in many ways (Geurs & Van Wee, 2004).

In Metronamica LUT, accessibility is defined as a composite measure of four components (RIKS, 2011). One of these, the local accessibility component, assigns a value to every land-use cell (in this research 250 x 250 m.) based on the distance to the closest transport element of some types. The local accessibility component is due for evaluation in this research.

The research can be divided in two parts. The first is an empirical analysis of land-use data of the Netherlands for the period 1989-2000. This data analysis analyses whether land-use developments are overrepresented along transport elements such as roads and railway stations and whether they prefer higher or lower quality levels of these elements (e.g. flow level). The second part uses this knowledge in the calibration of the local accessibility measure in Metronamica. It will be checked whether the simulation improves with a set of indicators that satisfies different aspects that should be evaluated in map comparison.

1.1 Goal

The focus of the research is on improving the modelling of transport system effects on land-use developments. Transport can influence location choice in different ways, such as by changing location accessibility and economic production in locations (see chapter 2). This research will focus on the impact via accessibility.

*The **goal of this research** is to improve the Metronamica LUT modelling of land-use changes due to transport system characteristics by providing empirical evidence and using this evidence to refine and calibrate the accessibility measure*

The proposed accessibility measure will be calibrated and validated for land-use maps of the Netherlands.

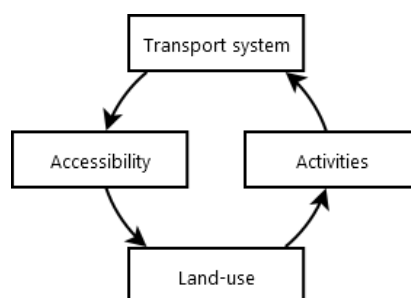


Figure 1.1 – Land-use transport feedback cycle (Wegener & Fürst, 1999)

1.2 Transport, accessibility and land-use

Transport; accessibility; land-use: these terms are not only important individually, but in this particular order as well. Accessibility in Metronamica is partly derived from the travel costs over roads and partly from the distance to any infrastructure element defined by the user (see section 3.5). For the application of the Netherlands, the latter includes only transport-related elements. Therefore, the following definition of a transport element is used:

*A **transport element** is one unit of any type of transport infrastructure, such as a train station, highway ramp and road segment. A **traffic characteristic** refers to a quantitative characteristic that stems from the traffic over the element, like flow (veh/h), congestion (flow/capacity ratio) and operating speed (km/h).*

*A **land-use type** is the occupation of a piece of land that comes forth from the activities that are undertaken.*

*A **land-use change** is a change in time of the land-use type that occupies an area. The more recent land-use type is considered the value of the change.*

The 'value' is thus the land-use type the area changed to rather than changed *from*. The cause of land-use changes lies in the location decision behaviour of actors involved in shaping land-use patterns, such as residents and companies. It should be kept in mind that land-use changes reflect the location decisions of a variety of actors.

Accessibility, then, is the means to understand and model at least part of the relationship between transport elements and land-use changes. Although accessibility is typically a broad, multi-interpretable concept, a general definition was given by Geurs & Van Wee (2004):

***Accessibility** is the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s).*

Chapter 2 will further describe literature about land-use transport interaction theories and the accessibility concept.

1.3 Research questions

Given the notion above that transport elements influence land-use developments in theory, the first question that arises is what influences of which transport elements on which land-use developments can be found in reality and to what extent. The following question is how accessibility modelling can be improved in Metronamica using the results from the first question. The third question would be whether this improvement affected the simulation results in a positive way.

Can an adjusted accessibility measure improve the modelling of land-use effects of transport network characteristics?

1. What is the role of transport elements and accessibility in explaining land-use changes in the Netherlands?
 - a. What method can be used to determine the influence of transport elements on location behaviour? Section 5.1
 - b. Are certain land-use changes over- or underrepresented in the proximity of certain transport elements? Section 5.3
2. How can accessibility modelling be improved in Metronamica LUT?
 - a. How is accessibility defined currently? Section 3.5
 - b. What suggestions for improvements can be made using the data analysis results from research question 1? Section 6.1

3. Do the suggestions from question 2 improve the simulation performance?
 - a. What indicators can be used to measure the performance of a land-use simulation? Section 6.2
 - b. Does the land-use simulation perform better after using the changes, according to the chosen indicators? Section 6.3

The data analysis and calibration focuses on the area of the Netherlands. Land-use occupation maps of the Netherlands are available for three years: 1989, 2000 and 2006; with a map resolution of 250 x 250 m. See chapter 4 for more information about the case study and available data.

*The **case study area** is defined as the Netherlands as delimited by the country's borders, excluding the area occupied by the North Sea and Wadden Sea.*

1.4 Report outline

Chapter 2 to chapter 4 serve as an overview of information, while chapter 5 and 0 contain the methodology and results of the research. Chapter 2 will first of all discuss relevant literature. This includes land-use transport interaction theories as well as results from empirical studies on the same subject. Furthermore, the term accessibility will be discussed in more detail. Chapter 3 follows with an overview of Metronamica LUT where the most relevant parts of the model are described in detail. The data for the Metronamica application of the Netherlands are presented in chapter 4. The first part of the research, the data analysis, is presented in chapter 5. The methodology, the results and a discussion of the latter are included. The second part of the research, the application in Metronamica, can be found in chapter 0. The findings after calibrating the local accessibility component, using the data analysis results, are described there.

2 Theory

It is not easy to determine how transport influences land-use. It is widely understood that transport infrastructure determines the accessibility of locations. In turn, accessibility is seen as an important indicator for location choices of different actors (Wegener & Fürst, 1999).

However, accessibility is not seen as the only intermediate variable between transport and land-use. Economic factors such as agglomeration effects and productivity in locations are described as intermediate factors as well (Bruinsma, 1994; De Bok, 2007). The importance of complementary policies, momentum and promotion of transport network changes have also been identified as important intermediate factors (Polzin, 1999).

Moreover, there are many other, non-transport related factors that influence the location choice of actors. While the influence of accessibility is mostly described as 'slight, but significant', it is always noted that other factors are more important in explaining location choice (Bruinsma, 1994; Iacono & Levinson, 2008; Zondag, 2007).

The complexity of location decision is therefore eminent. This chapter provides an overview of theories and empirical findings in order to give an overview of the relevant understandings in this research field. Section 2.1 provides some theoretical models about the interactive relationship between land-use and transport. Section 2.2 describes some of the empirical researches in this field. One of these empirical studies is highlighted in section 2.3, since it focuses on the Netherlands. Finally, section 2.4 gives a short overview of the considerations to be made when defining 'accessibility'.

2.1 Land-use transport interaction theories

It is often noted that much research has been done in order to explain the relation between land-use and transport, yet scattered among the different research fields of transport, geography and economy (Bruinsma, 1994; Zondag, 2007). The following is an overview of theoretical models that explain (part of) this relationship.

A review research on land-use transport interaction theories was done by Wegener & Fürst (1999). They distinguished three different theoretical approaches: technical theories (urban mobility systems), economic theories (where cities are seen as markets) and social theories (where societies reflect in urban space). What these models have in common is the notion that transport and land-use influence each other; a relation that was first studied by Hansen (1959) and resulted in the concept of the land-use transport feedback cycle in Figure 1.1. Starting at the top, the transport system determines how accessible locations are. Accessibility plays, in turn, a role in the location choices. The land-use patterns determine where people undertake their activities, resulting in traffic flows on which the transport system anticipates, etc.

The main interest of this research is at the relationship from transport to land-use (the left part of the feedback cycle). The three types of theories (technical, economic and social) all explain part of this relationship. A summary of the expected impacts of all three theory types was made by Wegener & Fürst and is given in Table 2.1.

In order to connect the scattered studies in the fields of transport, geography and economy, Bruinsma set up a model that integrates the scattered findings, see Figure 2.1 (Bruinsma, 1994). It can be seen that a certain infrastructure change influences the land-use (bottom right) in two ways. Both relations start with a change in transport costs due to the transport infrastructure change. The first 'path' is via location accessibility. This relation is similar to the left part of the land-use transport feedback cycle. The second path is an economic relationship. Bruinsma argues that a change in transport costs will lead to an adjusted productivity of households and companies. This change in productivity will lead to land-use changes as well. An important

Direction	Factor	Impact on	Expected impacts
Transport ↓ Land use	Accessibility	Residential location	Locations with better accessibility to work-places, shops, education and leisure facilities will be more attractive for residential development, have higher land prices and be developed faster. Improving accessibility locally will change the direction of new residential development, improving accessibility in the whole urban area will result in more dispersed residential development.
		Industrial location	Locations with better accessibility to motorways and railway freight terminals will be more attractive for industrial development and be developed faster. Improving accessibility locally will change the direction of new industrial development.
		Office location	Locations with better accessibility to airports, high-speed rail railway stations and motorways will be more attractive for office development, have higher land prices. Improving accessibility locally will change the direction of new office development.
		Retail location	Locations with better accessibility to customers and competing retail firms will be more attractive for retail development, have higher land prices and be faster developed. Improving accessibility locally will change the direction of new retail development.

Table 2.1 – Expected theoretical impacts of transport on land-use according to technical, economical and social theories (Wegener & Fürst, 1999)

note is that intermediate factors, such as developments in economy, demography, technology, environment and general government policy can influence the strength of all the indicated relationships.

Zondag constructed a more general model for the interaction between transport, economy and land-use on a regional level (Zondag, 2007). The similarities with the above land-use transport interaction models are clear. Firstly, the influence of transport on land-use takes place via the accessibility of locations. Secondly, similar to Bruinsma, transport reduces transport costs and therefore affects the regional economy, which, in turn, determines the land-use patterns. The size and kind of the relations in this model are affected by space, time and actors. A difference with Bruinsma is the notion that accessibility does not necessarily change due to reduced transport costs (another factor might be the addition of more modalities). Note that neither model defines the term accessibility, so it might be pre-occupied to immediately relate the effects exclusively to transport costs.

De Bok (2007) has researched transport effects specifically on firm dynamics and mentions two intermediate variables between transport and land-use, that are interrelated. One of them is location accessibility and the other, closely related to accessibility, is agglomeration. A distinction is made between localisation

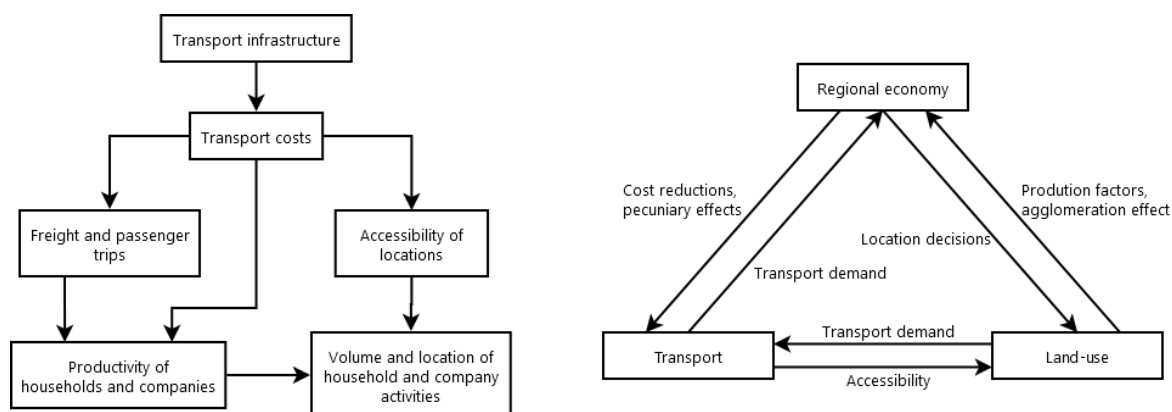


Figure 2.1 - Alternative models of land-use transport interaction. Left: integrated model of transport, land-use and economy (Bruinsma, 1994, freely translated). Right: transport, economy and land-use interaction model for a regional scale (Zondag, 2007).

agglomeration (i.e. area with many firms in same industry) and urbanisation agglomeration (i.e. high density urban area). Both types of agglomeration have significant advantage over other locations.

The development potential of an area is determined by its valuation by relevant stakeholders. Accessibility is one of factors in this valuation. As will be discussed in the next section, the role of accessibility in this valuation is more important for companies than for residents (Bruinsma, 1994; Zondag, 2007).

Note: in all of the models described above, the strength of the relations is highly influenced by external factors. Government regulations can strictly guide land-use development and can thus prevent the system from developing according to these models. This is especially the case in the Netherlands, where land-use developments are controlled by the (national) government to a relatively high extent. What needs to be taken into account, is that the effect of accessibility on job and resident distribution is moderate in free markets and negligible in fully restricted markets (Zondag, 2007). Other contextual factors are of great influence as well. For example, the absolute effect of transport changes is many times higher in developing countries than developed countries, which can be explained by the density of the current transport network.

Conclusions

The models described so far are only a selection of the attempts to structure the complexity that surrounds the transport land-use relationship. It became clear that for example productivity is a major factor in addition to accessibility and that the relationships are not straightforward at all. For this research, though, the focus lies on the interactions between the transport and the land-use via accessibility. Therefore, the land-use transport interaction feedback cycle is the most appropriate guideline for this research.

It was seen in Table 2.1 that the role of accessibility in attracting land-uses differs per land-use type. Therefore, it is important to make a clear distinction between land-uses and compare the results to each other, in order to get a good overview of the impacts of accessibility.

2.2 Empirical findings

Different researchers have noted that the impacts of transport on land-use have attracted much less research attention than the reverse effects of land-use on transport. Multiple reasons can be indicated for this lack of attention (Wegener & Fürst, 1999; Zondag, 2007):

1. Location accessibility hardly changes by a transport/land-use policy in developed countries, since the transport network is already very dense and developed. These effects would be larger in developing countries, but there has not been much attention for these areas either.
2. There is a large time-gap between transport policy implementation and the corresponding land-use effects (at least 10 years). As indicated by Wegener (2004), land-use is among the slowest changing components of a regional system.
3. Land-use changes are subject to many other influences than transport and direct transport impacts are therefore difficult to isolate. Examples of other factors are such as population growth, household formation, production technology.

These are indications that research to the determinants of land-use change is quite difficult and that most results yield significant uncertainty. Therefore, not many strong empirical relationships have been established (Polzin, 1999). With this in mind, the following is a short overview of empirical findings on the impacts of transport on location choice.

In their review study on empirical studies, Wegener & Fürst (1999) concluded the importance of accessibility varies among different land-use types. Residential areas with high accessibility are likely to be developed faster and the development is more dispersed when all areas have equal accessibility. In general, accessibility does not seem to be an important factor for industrial locations, except for high-tech and service firms. For retail and office land-use, however, accessibility is an important factor. Office developments are mainly seen at highly accessible locations in the inner-city or city peripheral with good motorway access. Development in retail location is found in highly accessible inner-city locations and peripheral locations with ample parking and good road accessibility.

Zondag (2007) concludes in a similar review that accessibility has a modest, but significant influence on residential location decision. Other variables are more important, which are demographic variables, neighbourhood characteristics and especially housing attributes. It is sometimes argued that housing prices are affected by accessibility (and thus transport), but there are different study results as to how large this influence actually is. The location choice of firms (including office, industry and retail) is in general more sensitive to accessibility and especially to highway accessibility, although accessibility is more important for relocating national firms than for immigrating foreign firms. Land price, available subsidies and location space are, however, more important than accessibility. For both residential and firm location choice, there is not much evidence that public transport influences location choice, except for specific industry sectors and areas with a high public transport share. Additionally, in residential location behaviour, accessibility is a significant variable in the decision to move (i.e. insufficient accessibility at current location), while it is not significant in the choice of the new location. The extent to which the population will grow when accessibility increases is larger than the extent to which the population will shrink when accessibility decreases.

The micro scale dynamics between transport infrastructure and firm dynamics specifically were studied by De Bok (2007). Six components of firm change were identified, including movement of a firm. The movement of a firm is divided in two phases. Phase 1 entails the decision to move, for which the size, age, growth rate and industry type of the firm appeared to be the most significant variables. The influence of accessibility is described as small. For agglomeration effects it was seen that firms in more diverse locations have a higher move probability. Phase 2 is about the location where to move to and a higher influence of transport is found. Note that these results are the exact opposite of the conclusions of Zondag for residential location choice. Additionally, location preferences differ among industry sectors and nearby locations have a higher preference as well. Mature firms prefer more accessible and specialised (i.e. higher agglomeration) locations, whereas young firms prefer diverse locations. Good labour market accessibility was specifically important for business services, manufacturing, transport warehousing & communications and trade & retail. The preference between road and rail accessibility differs between industry sectors as well. Companies generally tend to move to areas with localisation agglomeration as well.

Bruinsma conducted several empirical studies of which the main conclusions are the following (Bruinsma, 1994). Firstly, transport projects do not only change accessibility in the area of the project, but other areas as well. For city and metropolitan studies, it is found that accessibility benefits of transport changes are the greatest in already high accessible locations. Secondly, market perspective and internal factors are the most important factors to change the workforce (productivity) and therefore location choice. Transport can be labelled as one of the factors in the second important category. Thirdly, the evaluation of a certain location by entrepreneurs has a high correlation with location accessibility. Lastly, it appears in the literature overview that near-highway locations are generally highly valued by companies, while centre locations are not.

Levinson's researches show that residential and commercial developments in London followed the construction of subway lines and vice versa. (Levinson, 2008, 2011). For the New York subway, however, no convincing link to residential developments was found. It was also found that first subway stations in a network maintain the largest share of travellers over time (Levinson & Xie, 2011), indicating that land-use developments following transport developments. The other way around, it is also true that the first station is likely built in the busiest area.

Finally, Iacono & Levinson (2008) tried to establish the determinants of land-use patterns. Both transport and non-transport related determinants were investigated. The non-transport related factors appeared to be more significant in explaining land-use changes, although the inclusion of transport factors contributed positively to the overall goodness-of-fit of the model. The most striking transport land-use related correlation was found between commercial/industrial land-use types, and proximity to highways and high accessibility levels (measured as number of jobs weighted by access travel time).

Conclusions

The results from the empirical studies are summarised in Table 2.2. The second column is taken from Wegener & Fürst (1999). The third column adds results from the other discussed studies (of which the most are more recent). When taking all the information into account, a complete picture of the transport land-use

relationship does not emerge. Rather, the information is composed of loose observations that, when taken together, do not lead to a complete, consistent picture. Besides, the information is taken from different studies with different objectives, conditions and study areas, so extra caution must be taken in comparing the information. Furthermore, almost all studies note that other factors contain more explanatory power for land-use changes than accessibility, although accessibility is always mentioned as a variable of 'second class' importance.

The data analysis in this research aims to contribute to the empirical findings by focussing on the influence of transport elements (roads, ramps and railway stations) on land-use developments in the Netherlands.

2.3 Empirical findings for the Netherlands

The magnitude to which determinants of land-use developments are important can differ greatly among regions (Bürgi & Turner, 2002). Verburg et al. (2004) researched land-use determinants specifically for the Netherlands. They identified different theoretical disciplines which all provide factors that potentially influence land-use patterns. A selection of these factors is used in a logit estimation analysis to determine which of these are important. The identified disciplines are biophysical constraints and potentials, economic factors, social factors, spatial interaction/neighbourhood characteristics and spatial policies.

Vervurg et al. examined what factors determined the land-use pattern of 1989. A similar analysis was done for the period 1989-1996. A number of soil variables and altitude show a significant contribution to the explanation of the spatial patterns of nature and arable land for 1989. The explanation of residential and industrial land-use patterns is very poor. This implies that other factors, that were not included in this analysis, are more important. For the recent time period, variables that are significant for new residential areas relate to the biophysical conditions (drier locations, lower reclamation costs), accessibility and policy. For new industrial/commercial developments only accessibility and location near highways is of importance and for new recreational areas the location close to concentrations of population and the growth-centre policy.

Location choice	Wegener & Fürst	Other studies	Factors identified as more important in location decision
Residential	More accessible locations are developed faster. If accessibility in the whole region grows, residential development will be more dispersed.	Influence is "modest, but significant". Accessibility has significant influence on the decision to move, but no significant influence on the choice of the new location.	-Demographic characteristics -Neighbourhood characteristics -Housing properties
Industrial	There is little evidence of impacts of accessibility on location of manufacturing, but ample evidence of the importance of accessibility for high-tech and service firms.	Highway accessibility is important. There is no evidence that PT accessibility has any influence.	-Land price -Available subsidies -Space -Market perspective -Internal factors
Office	Office development occurs predominantly at highly accessible inner-city locations or in office parks or 'edge cities' at the urban periphery with good motorway access.	In contrary to residential choice, accessibility has no proven significance in move choice, yet is significant in new location choice.	
Retail	Retail development occurs either at highly accessible inner-city locations or on peripheral sites with ample parking and good road accessibility.	Firms move to locations with more location agglomeration. Labour market access is in particular important for business services, manufacturing, transport warehousing and retail	

Table 2.2 – Overview of empirical findings on the influence of transport on land-use

The studies were additionally performed again for three selected regions in the Netherlands. The results suggest that spatial disaggregation leads to more specific relation between the location characteristics and land-use. For the recent time period, accessibility was important for all regions, yet each region 'preferred' another accessibility measure.

The estimation for the recent time period was finally done once more, this time including neighbourhood characteristics. The results show that the overall explanation increases, although not all neighbourhood factors are significant. The positive autocorrelation between residential and industrial land-uses indicates that urban growth is more important than new development outside current cities. New recreation areas are associated with nearby forest/nature.

Concluding, historical agricultural and nature patterns show the strongest statistical relationship with soil and altitude. These biophysical factors are no longer the most important in recent land-use changes. Accessibility, spatial policies and neighbourhood interactions are more important. Since these factors are somewhat related to historical developments in land-use patterns, the authors conclude that "recent developments follow a process of self-organisation in which centripetal forces lead to the growth of existing residential centres." Additionally, the results indicate that policies have an important influence on land-use patterns in the Netherlands.

Furthermore, the compact urban development policy (1970-2000) of the Dutch national government was evaluated by Geurs & Van Wee (2006). Urban sprawl, car use, emission, noise and wildlife fragmentation levels would all have been higher without this policy. This is another indication that policy can, especially in the Netherlands, have a decisive influence on land-use developments.

Perhaps the most important conclusion is that land-use patterns cannot be explained by a single of the before mentioned disciplinary theories. Multiple factors from different theories are needed to adequately explain land-use change. Verburg et al. also issue a warning that found correlations do not directly translate into any causal relationships.

Conclusions

The research of Verburg et al shows that accessibility has increased in importance for explaining land-use patterns in the Netherlands in the last decades. It is one of three factors that show the most explanatory power, besides neighbourhood interactions and policies. Most importantly, the research supports the idea that accessibility is an important explaining variable for land-use patterns in the Netherlands. Studies for other areas show comparable results (e.g. Iacono & Levinson, 2008). However, in this research we are interested in expanding the understanding of the influence of accessibility by explicitly distinguishing effects of different transport element on different land-use developments.

2.4 Accessibility

Accessibility has been described in many different ways. For example, Morris (1978) denoted that accessibility reflects the *"ease with which activities can be reached from a given location using a particular transportation system"*. This and many other definitions contain the basic elements that were firstly denoted by Hansen as *"the potential of opportunities for interaction"* and more specifically as *"a measurement of the spatial distribution of activities about a point, adjusted for the ability and the desire of people or firms to overcome spatial separation"* (Hansen, 1959). The common element in these and other definitions is that accessibility is determined by both the reachable activities in destinations (determined by land-use patterns) and the resistance to travel in between (determined mainly by the transport system).

Geurs & Van Wee (2004) used the following definition for accessibility, which puts slightly more emphasis on the notion that the combination of land-use and transport systems define accessibility. This definition will therefore be used in this research as well:

Accessibility is the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s).

Still, this definition remains quite abstract. To make accessibility a useful concept for modellers and policy makers, it needs to be made operational (Borzacchiello, Nijkamp, & Koomen, 2010). Concrete measures are used to give an explicit value to it that differs spatially. Accessibility measures for the impact of land-use and transport developments and policy plans on the functioning of the society in general can be categorised into four groups (Geurs & Van Wee, 2004). These measures can be used for modelling purposes as well.

- **Infrastructure-based measures** analyse the performance of the transport network. Examples: travel times, congestion levels, network operating speed.
- **Location-based measures** analyse accessibility at different locations. Example: number of jobs within 30 minutes travel time. A further categorisation is made:
 - **Distance measures and contour measures:** a distance measure (or relative accessibility) is the degree to which two places are connected. Examples: straight line, average travel time, average speed. A contour measures analyses more than two places and Counts the number of opportunities which can be reached within a given travel time, distance or cost (fixed costs) or measures the time or cost required to access a fixed number of opportunities (fixed opportunities).
 - **Potential measure:** estimates the accessibility of opportunities in one zone to all other zones in which smaller and/or more distance opportunities provide diminishing influences.
 - **Adapted potential measures:** these add competition effects to potential accessibility measures. Different adapted potential accessibility measures can be distinguished.
 - **Balancing factors:** one of the adapted potential accessibility measures. The balancing factors ensure that the flow between two locations equals the activities in both origin and destination. The balancing factors can be seen as the accessibility measure.
- **Person-based measures** analyse accessibility at an individual level. Example: activities in which an individual can participate at a given time, time-space prisms.
- **Utility-based measures** analyse the economic benefit that is gained from accessibility to activities. These measures interpret accessibility as the outcome of a set of transport choices (i.e. people choose the transport option with the highest utility). A further categorisation is made:
 - **Logsum benefit measure:** describes the maximum expected utility of the full choice set.
 - **Balancing factors benefit measure:** calculates the benefits from each trip. The results should be similar to the logsum measure.
 - **Space-time measure:** includes available time in utility function of logsum.

It must be noted that no single accessibility measure will be perfect for a given problem. A combination of accessibility measures can improve the overall usefulness. Usually, a trade-off has to be made between scientific quality (theoretical components) and usability (communication, operationalisation). Metronamica's accessibility calculation also consists of four components, which will be discussed further in the next chapter. Bruinsma (1994) noted that simple indicators, like distance to the nearest transport element, might provide at least equal explaining power than more complex measures.

Conclusions

It can be concluded that there are a lot of possibilities in defining a measure for accessibility. A set of measures is usually the most useful. The available data for this research (chapter 4) are most suited for research using the distance to the nearest transport element as an accessibility measure. Moreover, as mentioned before, this relative simple measure is believed to contain at least equal explaining power for land-use developments than more complex measures. That is why the research focuses on this component (note that more components are included in Metronamica LUT, chapter 3).

2.5 Conclusions

Land-use developments are driven by a combination of determinants of which transport is one. The relation between transport and land-use can be described as interactive (Wegener & Fürst, 1999): the transport system determines the accessibility of areas, which in turn plays a role in land-use developments. The latter determines where activities take place, to which the transport system will be adapted.

It is noted by several researchers that in the empirical field, more research has been done on the impact of land-use on the transport system than the other way around. This research will focus on the influence of

transport on land-use developments through accessibility, in the first place to contribute the empirical understandings in the land-use research field and in the second place to improve the land-use development modelling in Metronamica.

There is no single best definition of accessibility. There are also many different indicators to quantify accessibility, which vary from simple to complex. This research focuses on the distance to nearby elements, perhaps the most simple accessibility indicator one can think of. However, it is noted that the distance to a nearby transport element might still be an equally strong explanation than more complex measures (Bruinsma, 1994). Nonetheless, a combination of measures is usually of greatest use (Geurs & Van Wee, 2004). Metronamica indeed uses a combination of accessibility measure, as will be described in the next chapter.

For the Netherlands specifically, a number of drivers can be identified as important explaining variables for land-use patterns. It was found that accessibility has increased in relative importance as an explaining variable over the last century (Verburg, Ritsema van Eck, et al., 2004). The methodology used in this research makes a further distinction between different transport elements (and different quality levels thereof). The data analysis results also provide information about the attraction of transport element to distance.

2.6 Implications for Metronamica LUT

It is clear that many aspects are important in determining land-use developments. These should be included in a land-use development model such as Metronamica LUT. Metronamica uses several sub-models to include the important factors that have been mentioned in this chapter. For example, economic productivity per region (section 2.1, Bruinsma) is included in the 'region interaction model'. Neighbourhood influences (section 2.3, Verburg et al.) can be set in the 'neighbourhood dynamics model'.

This research focuses on accessibility though, so these particular aspects are not due for evaluation. The accessibility model of Metronamica consists of several measures as well. One of those is a potential measure mentioned in section 2.4 (Geurs & Van Wee). Another measure, which will be the focus of the research, is the distance to the nearest transport element, called the 'local accessibility component'. The next chapter will describe Metronamica LUT in more detail.

3 Metronamica LUT description

The aim of this chapter is to provide a short overview of the Metronamica LUT land-use model and to give a deeper insight in how transport, land-use and accessibility are currently modelled. The information in this chapter is taken from Van Delden & Hurkens (2011), RIKS (2011) and Van Delden (2010), unless indicated otherwise.

3.1 Introduction to Metronamica LUT

Metronamica LUT is a spatial decision support system. These are systems that ‘support a decision research process for complex spatial problems’ (Densham, 1991). It is used by policy makers to evaluate the effects of policy alternatives on land-use developments.

Geonamica is the framework for Metronamica LUT and for other spatial models developed by the Research Institute of Knowledge Systems (RIKS). Geonamica is a software environment that does not include any models. For every application one or more models is taken from a model library. Metronamica LUT includes three models from this library: a cellular automata land use model, a four-stage transport model and a regional interaction model (Figure 3.1). The additional spatial indicator component is able to calculate some indicators for land-use maps, but indicator results do not impact on the main model components.

The models operate at either of two geographical levels: the regional or local level. Both the regional interaction and transport model operate at the regional level. The first generates and (re)locates activities (e.g. jobs, residents) among the regions. The second assigns traffic over the transport network, using transport zones which are typically smaller than the regions in the regional interaction model. The land-use model operates at a local level, meaning that it assigns land-use types to cells considering one region at a time (using the output from the other two models).

The regional interaction model provides employment and population totals for each region. It calculates the activity (number of people/jobs) for each economic sector in every region, the migration (number of people/jobs) between regions for each economic sector and the cell demand (number of cells) for each land-use function and region. This information is passed on to the land-use and transport models. The transport and land-use models are more relevant for this research and will be discussed in more detail in the following sections.

Note that the other Metronamica configurations, Metronamica SL and Metronamica ML, do not include the transport model. This research is carried out with Metronamica LUT and suggestions for the improvement of the model concern Metronamica LUT specifically. From here on, ‘Metronamica’ refers to Metronamica LUT.

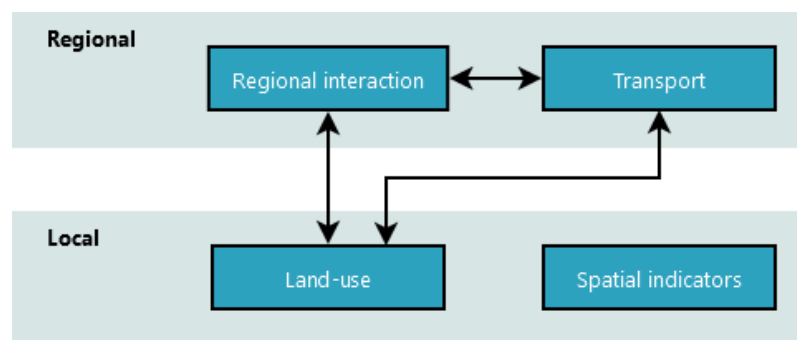


Figure 3.1 - Metronamica's sub-models and their corresponding geographical levels (RIKS, 2011)

3.2 Transport model

The demand for local travel and the traffic characteristics are determined in this model for private vehicles. The transport model basically follows the classical four-step transport model of generation, distribution, modal split and assignment (see e.g. Ortúzar & Willumsen, 2001) as illustrated in Figure 3.2. The land-use and regional interaction models provide input for the transport model and use its output as well. This is in line with the general understanding that there is an interaction between transport and land-use (see section 2.1).

The generation (or production-attraction) step calculates the number of produced trips in each origin and the number of attracted trips in each destination using the number of activities in each transport zone (from the activities in the larger zones of the regional interaction model). The distribution step couples produced and attracted trips, which results in a number of trips for every origin-destination pair. The modal split step decides which trip is made by what modality, by looking at the travel costs of each modality and at past behaviour. Finally, the assignment step assigns each trip to the route over the network with the least costs (composite measure of e.g. fuel use, travel time, distance, parking costs and tolls). The trip assignment leads to Wardrop's first equilibrium, where the leading rule is that no single user is able to choose a cheaper route (Ortúzar & Willumsen, 2001).

The output of the four-step model includes the operating speed, flow and congestion level on each link, as well as the generalised costs, generalised distance and generalised travel time for each origin-destination pair.

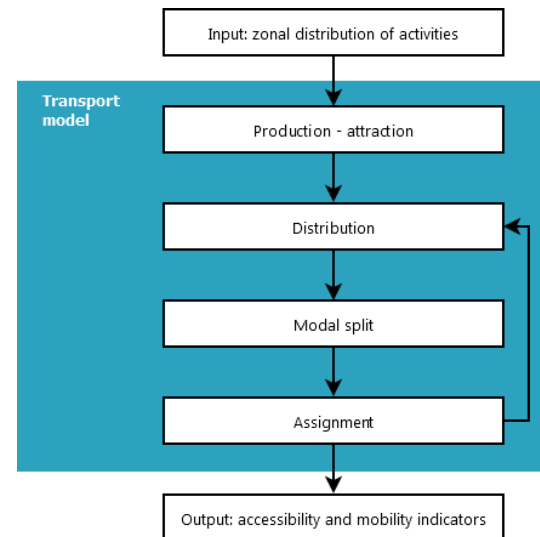


Figure 3.2 – Transport model flow chart (RIKS, 2011)

3.3 Land-use model

A cellular automata (CA) model uses a cellular grid in which each cell has a certain state (an attributed land-use type) that depends on the previous state of the cells within a certain neighbourhood of the cell (White, et al., 1997). Transition rules are set up to determine whether the cell's state should be changed, and if so, what the new state will be. In land-use modelling, the transitions rules are often expanded with other factors to account for e.g. accessibility effects and physical restrictions. CA models have been applied in land-use modelling for both urban (White, et al., 1997) and natural areas (Wickramasuriya, Bregt, Van Delden, & Hagen-Zanker, 2009).

The exact locations of the land-use types are determined on the local level with the cellular automata model. All cells in one region are assigned a land-use type, before considering the cells in the next region. A set of transition potential values is calculated per cell, one value for each potential land-use type the cell might change to. The cell then changes to the land-use function for which it has the highest transition potential until demands from the regional model are met. The changes take place in discrete time steps in which all cells change simultaneously. The transition potential for a certain cell and land-use type is determined by four drivers (Figure 3.3):

1. **Suitability** is the degree to which a cell can support a land-use function and is given in one map per land-use type. Each cell has a value indicating the suitability of the cell for the land-use type, which is composed of different factors such as flooding frequency, elevation and slope.
2. **Zoning** policies are also given in one map per land-use type. The maps are based on planning documents from any authority (e.g. specifying which areas are protected by law). The difference between suitability and zoning lies in the natural characteristics of the first and the manmade of the second.

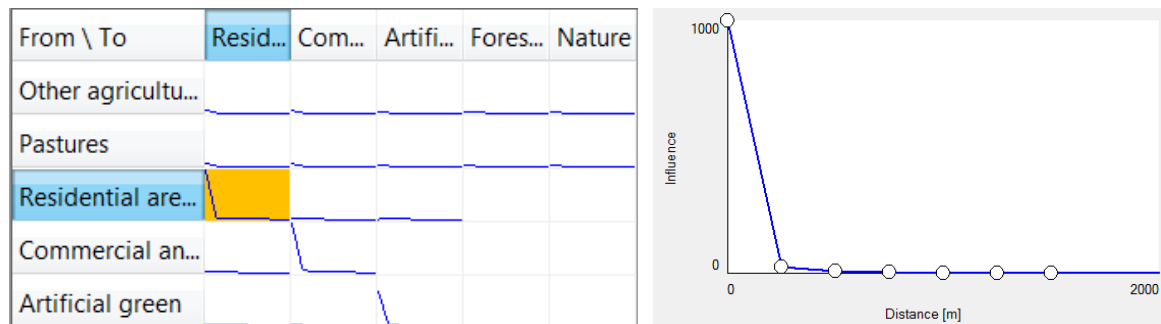


Figure 3.4 –Neighbourhood rules for the Netherlands application (left) and an enlarged view of the residential to residential graph (right)

It is important to realise that the overrepresentation graphs cannot be directly converted into Metronamica's attraction graphs for two reasons (Verburg, Ritsema van Eck, et al., 2004). The first is that the time-spans do not match: the statistical analysis usually entails a period of one or more decades, while Metronamica uses its neighbourhood rules for simulation time steps of one year at a time. The second reason is that the overrepresentation graphs do not prove any form of direct causality between land-use development, so modelling it in Metronamica involves making the assumption that attraction is the 'best' way to model certain developments.

The data analysis in this research will use a methodology analogous to the overrepresentation graphs of Verburg et al., but it will consider the neighbourhood around transport elements instead of cells. The resulting overrepresentation graphs can then be used to calibrate the local accessibility measure.

3.5 Accessibility

The accessibility driver is in fact a composite measure consisting of four components. As noted by Geurs & Van Wee, a combination of measures is often required to give a better understanding of accessibility in a certain area (Geurs & Van Wee, 2004). As can be seen in Figure 3.3, the zonal *accessibility* is determined in the transport model and the other components in the land-use model.

- **Zonal accessibility** is a potential accessibility measure (Geurs & Van Wee, 2004) that expresses the number of activities within reach corrected for transport costs to reach these activities.
- **Local accessibility** accounts for the distance between a cell and the nearest transport element. Each different land-use type has a distinct preference in how close or far it wants to be from certain links. The local accessibility is a measure for how much this preference is fulfilled.
- **Implicit accessibility** takes the current land-use of the cell into account. The cell will be more or less accessible for a land-use type when it is currently built-up compared to when it is currently not. The implicit accessibility's main goal is to include how developed the transport network within the cell already is.
- **Explicit accessibility** accounts for the fact that some land-use types cannot be crossed by activities that are generated by other land-use types.

If the considered cell is impassable, the total accessibility is equal to the explicit accessibility. In all other cases, the zonal, local and implicit accessibility are multiplied. The consequence of this multiplication (instead of a summation) is that only cells with high values for all three components are able to receive a high total accessibility. The reason for this is that all components are required to make a cell attractive. For example, consider the residential accessibility for a cell on a small island with a poor connection to the main land. It would be reasonable to assume the accessibility should be near the minimum. However, the local accessibility might be high since there is a road network on the island. The connection to the main road network is very poor, resulting in a poor value for the zonal accessibility. Not fulfilling the latter should thus lower the overall accessibility value significantly, regardless of the values of the other components. This example illustrates why the requirements of all components should be met to be able to truly say an area is accessible.

$$A_{f,c} = \begin{cases} EA_{f,c} & \text{if } f(c) \in LU_I \\ ZA_{f,c} * LA_{f,c} * IA_{f,c} & \text{otherwise} \end{cases} \quad 3.3$$

Where:

$A_{f,c}$	Accessibility in cell c for land-use f
$EA_{f,c}$	Explicit accessibility in cell c for land-use f
$ZA_{f,c}$	Zonal accessibility in cell c for land-use f
$LA_{f,c}$	Local accessibility in cell c for land-use f
$IA_{f,c}$	Implicit accessibility in cell c for land-use f
LU_I	Set of impassable land-use types (parameter)

3.5.1 Zonal accessibility

The zonal activity is the only component that is calculated on a transport zone level rather than a cellular level (i.e. every cell in a transport zone has the same value). It is a potential accessibility measure that can be described as the number of activities within reach corrected with a distance decay factor. An activity a can be jobs, residents or number of cells of a certain land-use type. The zonal activity in zone z for activity a is calculated as:

$$ZA_{z,a} = \sum_{z'} TA_{z',a} e^{-c_{avg,z,z'} * \gamma_a} \quad 3.4$$

$$c_{avg,z,z'} = \frac{c_{z,z'} + c_{z',z}}{2} \quad 3.5$$

Where

$ZA_{z,a}$	Zonal accessibility for transport activity a in zone z
$TA_{z',a}$	Level of transport activity a in zone z' (input from regional interaction model)
$c_{avg,z,z'}$	Average transport costs between zone z and z'
γ_a	Costs sensitivity parameter for activity a (parameter)
$c_{z,z'}$	Transport costs from z to z'

The zonal activity in zone z and any land-use function f is then calculated as follows. The values are normalised in the interval $[ZA_{min}; 1]$. The minimal zonal accessibility can be interpreted as a relative weight of the zonal accessibility compared to the other components, where a lower value equals a greater importance.

$$ZA_{f,z} = ZA_{min} + (1 - ZA_{min}) \sum_a \frac{FA_{f,a} * ZA_{z,a}}{ZA_{max,a}} \quad 3.6$$

$$ZA_{max,a} = \max_z (1; \max_z (ZA_{z,a})) \quad 3.7$$

Where:

$ZA_{f,z}$	Zonal accessibility of land-use type f in zone z
ZA_{min}	Minimal zonal accessibility (parameter)
$FA_{f,a}$	Proportion of cell with land-use f that is considered in activity a

3.5.2 Local accessibility

The local accessibility is calculated on a cellular level. It is based on the distance between a cell and the nearest transport links of each link type. First of all, the accessibility in cell c to link type s for function f is calculated as follows:

$$LA_{s,f,c} = \begin{cases} \frac{a_{s,f}}{D_{s,c} + a_{s,f}} & \text{if } a_{s,f} > 0 \\ 0 & \text{if } a_{s,f} = 0 \\ 1 - \frac{|a_{s,f}|}{D_{s,c} + |a_{s,f}|} & \text{otherwise} \end{cases} \quad 3.8$$

Where:

- $LA_{s,f,c}$ local accessibility of cell c to link type s
- s link type
- c cell
- f land-use function of cell c
- $D_{s,c}$ distance between c and nearest cell that is covered by link type s
- $a_{s,f}$ distance decay (parameter)

This formula can be explained as follows. A positive distance decay value means that land-use type prefers to be closer to the transport element. The lower the value, the closer it needs to be in order to fulfil its preference, so the value of $LA_{s,f,c}$ decreases more rapidly when the distance becomes higher. The other way around, a negative distance decay value means that the land-use type prefers not to be close to the transport element. The relation between the distance decay and $LA_{s,f,c}$ are seen in Figure 3.5.

The next step is to aggregate the values over s , starting with the total accessibility for all link types with negative distance decay parameters.

$$LA_{f,c}^{neg} = \prod_{s \in S_f^-} w_{s,f} * LA_{s,f,c} \quad 3.9$$

$$w_f^{neg} = \prod_{s \in S_f^-} w_{s,f} \quad 3.10$$

Where:

- $LA_{f,c}^{neg}$ Total negative local accessibility of cell c for land-use f
- S_f^- Set of all link types with negative distance decay parameter for land-use f
- $w_{s,f}$ Relative weight of the proximity to different networks on total local accessibilities $\in (0; 1)$ (parameter)
- w_f^{neg} Total weight of negative local accessibility

This is then aggregated with all links with a positive distance decay parameter as shown below. This formula works by subtracting a certain factor from 1 in the nominator. The subtraction becomes smaller as more values of $LA_{s,f,c}$ are present and when the individual values are higher. As a result, the local accessibility for the cell increases. This is in accordance with the idea that when one road cannot fully fulfil a cell's need to be accessible, another road can fulfil the remaining need.

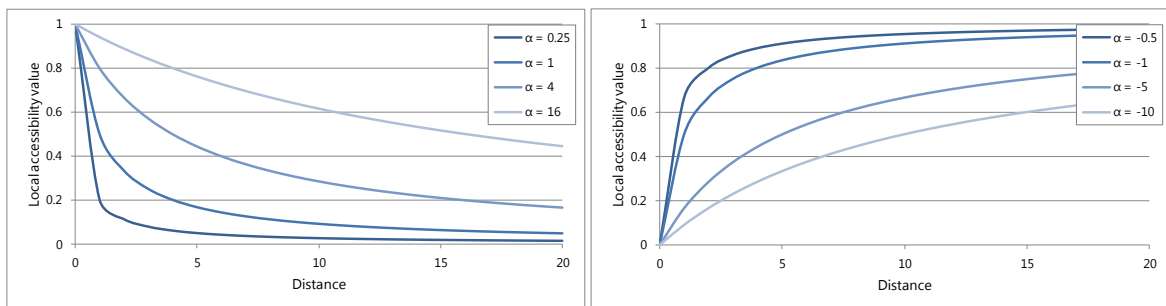


Figure 3.5 – Influence of distance decay parameter on the cell's local accessibility (left: positive values; right: negative values)

$$LA_{f,c} = \frac{1 - (1 - w_f^{neg} * LA_{f,c}^{neg}) * \prod_{s \in S_f^+} (1 - w_{s,f} * LA_{s,f,c})}{1 - (1 - w_f^{neg}) * \prod_{s \in S_f^+} (1 - w_{s,f})} \quad 3.11$$

Where:

$LA_{f,c}$ Local accessibility of cell c with land-use function f

S_f^+ Set of all link types with a positive distance decay parameter for land-use f

3.5.3 Implicit accessibility

When a cell is occupied by a certain land-use type, assumptions can be made for the way the traffic network is going to be developed within the cell. Metronamica does not model this development directly, but the implicit accessibility incorporates this element. The cell's current function is crucial: when the current state is urban, the cell is probably more accessible or has at least a developed transport network to a certain degree. This difference can make the cell more or less attractive for other land-use types to take over. For example, the physical presence of a transport network makes currently built-up areas more attractive for urban than for non-urban developments.

$$IA_{f,c} = \begin{cases} Urb_f & \text{if } f(c) \in LU_U \\ N Urb_f & \text{otherwise} \end{cases} \quad 3.12$$

Where:

Urb_f Implicit accessibility for land-use f of a cell that is occupied by an urban land-use (parameter)

$N Urb_f$ Implicit accessibility for land-use f of a cell that is not occupied by an urban land-use (parameter)

LU_U Set of urbanised land-uses (parameter)

3.5.4 Explicit accessibility

Finally, the explicit accessibility accounts for the fact that some land-use types are impassable for activities generated by other land-use types (but not for activities generated by the same land-use type). When the cell's current land-use type is flagged as impassable, the total accessibility is set to the value of the explicit accessibility (formula 3.3).

$$EA_{f,c} = \begin{cases} IA_{f,c} & \text{if } f(c) = f \\ 0 & \text{otherwise} \end{cases} \quad 3.13$$

Where:

$f(c)$ Land-use occupied by cell c

3.6 Metronamica in this research

This chapter gave an overview of the four drivers that are used to model land-use developments in Metronamica. Accessibility is one of those drivers and is composed out of four different components. The local accessibility component is the focus of this research. This is a measure for the distance of a cell to the closest transport element. The distance decay and weight parameters will play an important role in chapter 0. Nonetheless, local accessibility is but one of the many variables in Metronamica and it should not be forgotten that it determines land-use changes in conjunction with many other factors.

4 Data

The data analysis and calibration are applied to data from the Netherlands. Land-use maps for the years 1989, 2000 and 2006 are available and will be used to generate land-use change maps. These are used for the data analysis. The available data on transport consist of railway stations, highway ramps and a vast road network (including highways to 30 km/h roads). The road network is reclassified to make a more appropriate and consistent classification for the data analysis and the application in Metronamica. This will all be described in detail below.

4.1 Sources

The data used in this research consist of three land-use maps, several transport maps and a region map. These are taken from two sources (Table 4.1): a Metronamica project file made available by RIKS and Rijkswaterstaat Dataportal (Rijkswaterstaat, n.d.). The Metronamica data are adjusted data from Corine (European Environment Agency, n.d.) and Rijkswaterstaat, but are seen as the 'unaltered' input data for this research. Technical information about the datasets can be found in appendix A.

4.2 Land-use data

The land-use data are given in raster maps with a cell size of 250 x 250 m. Three historical land-use maps are available for different years: 1989, 2000 and 2006. Figure 4.1 shows the land-use maps which are, for visualisation purposes, already corrected for the study area (i.e. land-use occupations outside the study area are not displayed).

4.2.1 Land-use classification

The land-use maps contain fourteen land-use types. This particular classification is the result of a reclassification process which had been carried out on the original Corine maps (some classes were already merged together). The original classification and reclassification process were the same for all three maps. Thus, the noise between the years is limited in this regard. Nonetheless, it should not be forgotten that noise between the maps is inevitable, e.g. due to errors in the map making process.

Data	Source	Used in Metronamica	Year
Metronamica road map - Road type - Simulated flows - Simulated congestion - Simulated operating speed	Rijkswaterstaat (NWB)	Yes	2004
Rijkswaterstaat road map Observed flows	Rijkswaterstaat (NWB)	No	1998
Train station map Station type	Rijkswaterstaat	Yes	2004
Ramps map	Rijkswaterstaat	Yes	2004
Regions map	Rijkswaterstaat	Yes	
Adjusted Corine land-use map	Corine	Yes	1989
Adjusted Corine land-use map	Corine	Yes	2000
Adjusted Corine land-use map	Corine	Yes	2006

Table 4.1 – Data used in this research and its sources. The bullet points show some relevant (but not all) attributes of the dataset that were used in this research.

Land-use categories	Used in research	Modelled as	Reason for exclusion
Other agricultural areas	Yes	Vacant	-
Pastures	Yes	Vacant	-
Residential areas	Yes	Function	-
Commercial and industrial areas	Yes	Function	-
Artificial green	Yes	Function	-
Forests	Yes	Function	-
Nature	Yes	Function	-
Wetlands	No	Feature	Most important part outside study area (Wadden Sea)
Fresh water	No	Feature	Very few cells in total, very few cells in change map
Marine water	No	Feature	Mostly outside study area
Infrastructure	No	Feature	Very few cells in total, very few cells in change map
Mines and dumpsites	No	Feature	Very few cells in total, very few cells in change map
Port areas	No	Feature	Very few cells in total, very few cells in change map
Airports	No	Feature	Very few cells in total, very few cells in change map

Table 4.2 – Overview of land-use types that were and were not taken into account in this research

Seven out of the fourteen classes were found relevant for this research (Table 4.2). To understand the exclusion of the other land-use types, the characteristics of the land-use maps and the land-use change maps have to be analysed. Figure 4.3 gives the occupation (expressed in the number of cells) of each land-use type in 1989, 2000 and 2006. Ultimately, three main reasons can be identified for excluding wetlands, fresh water, marine water, infrastructure, mines & dumpsites, port areas and airports:

- The upper five land-use types in Figure 4.3 each cover less than 0.5% of the study areas. Similarly, they occupy less than 1% of all cells in the change map. The results of the data analysis will be less reliable for these types since only few instances of new cells are available. (Note that the same can be said for pastures, although the difference is that pastures still occupy a large area of the Netherlands despite of the few changes).
- The excluded land-use types are typically modelled as 'features' in Metronamica, meaning that the total number of cells and their locations are fixed. Nonetheless, information about the behaviour of these land-use types with respect to accessibility might be interesting in some cases, but this information would be less relevant.
- Although the wetlands are spread through the Netherlands, the largest part is the area occupied by the Wadden Sea. Similarly, marine water cells largely fall outside the study area as well.

4.2.2 Land-use change maps

Land-use change maps were created using the land-use maps (Figure 4.2). With the availability of three land-use maps, two periods can be established: period A (1989-2000) and period B (2000-2006). One land-use change map was created for each period, reflecting the changes within this time span. The creation was done with the following rules:

- When a raster cell has the same value in both land-use maps, the cell in the land-use change map will not contain any data.
- When a raster cell has a different value in each of the land-use maps, the value in the land-use change map is set to the value in the most recent land-use map.

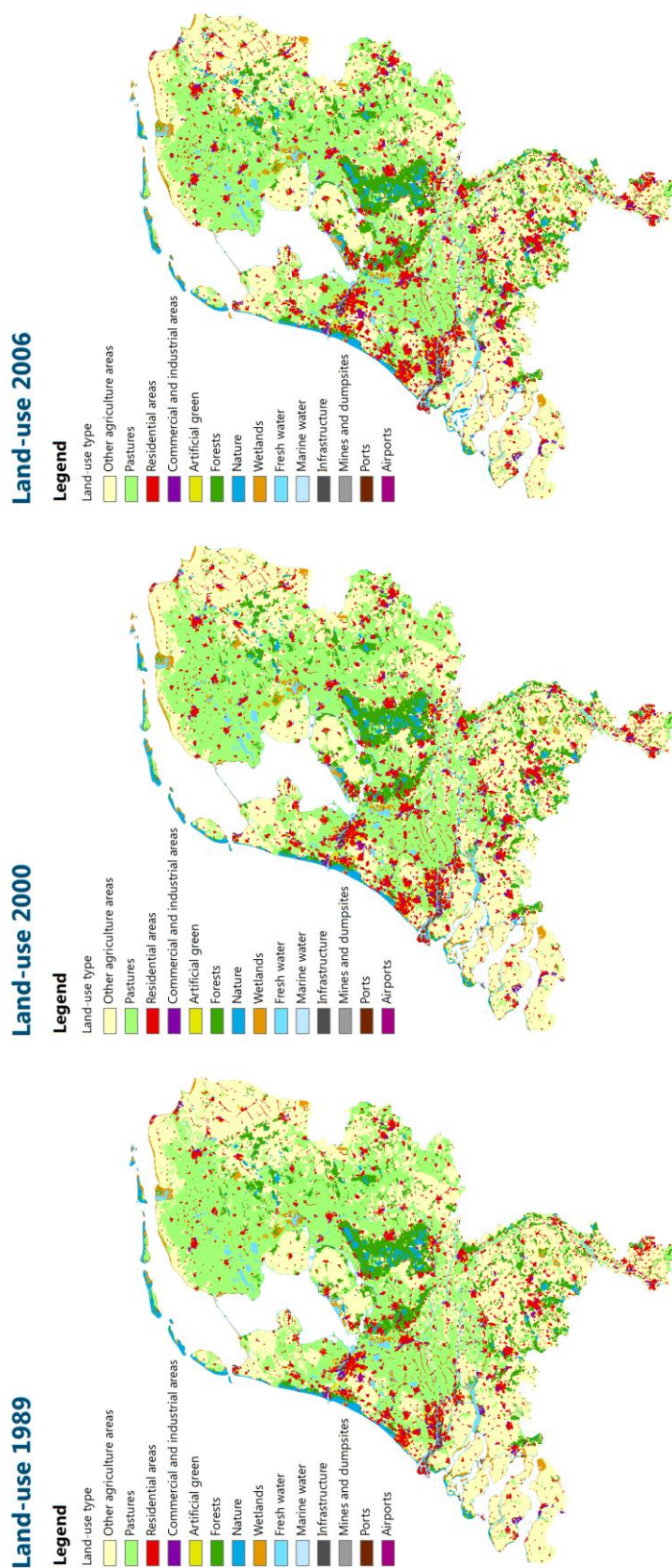


Figure 4.1 – Available land-use maps for the Netherlands

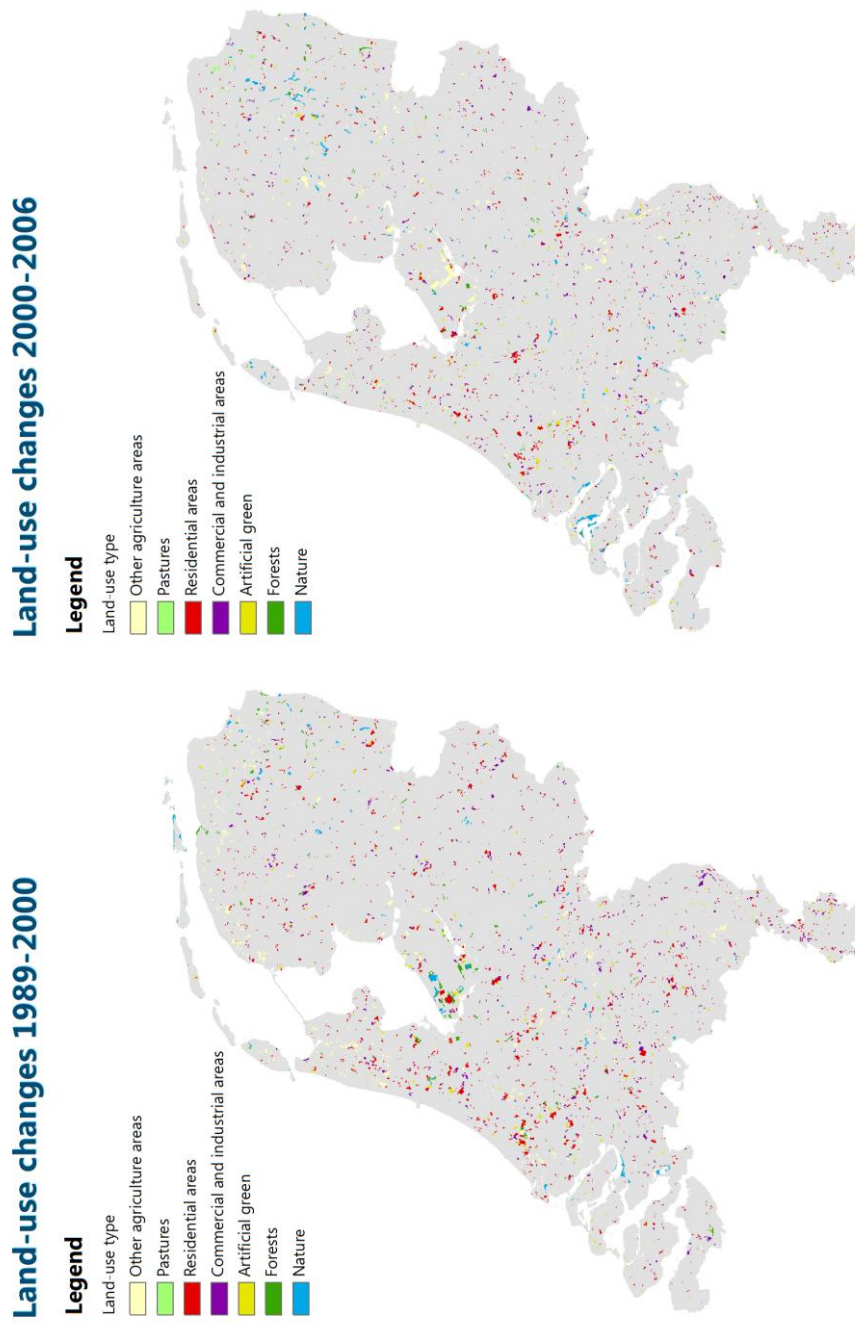


Figure 4.2 – Land-use change maps

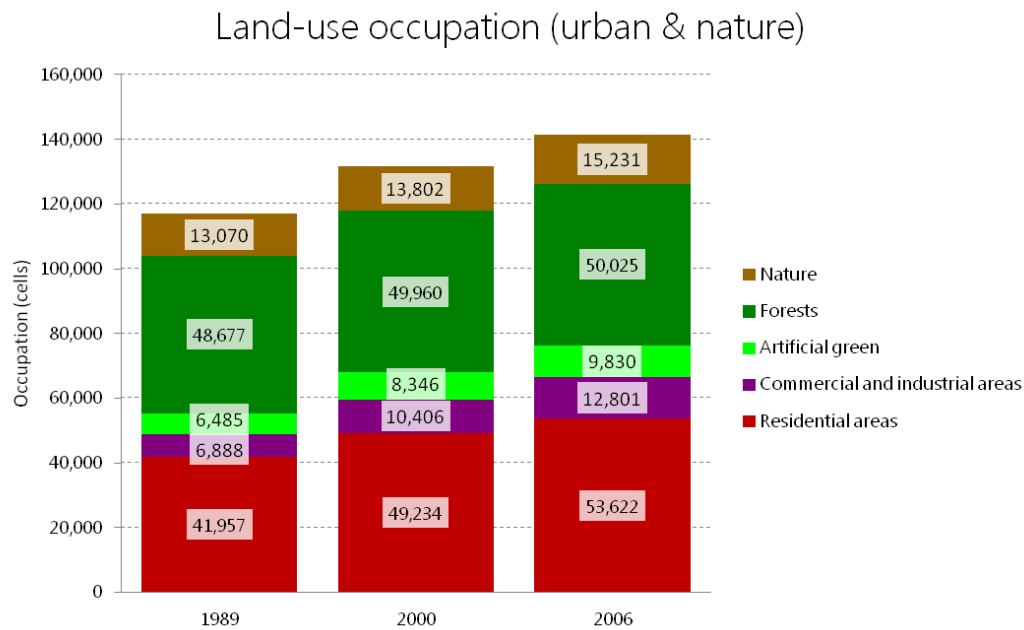
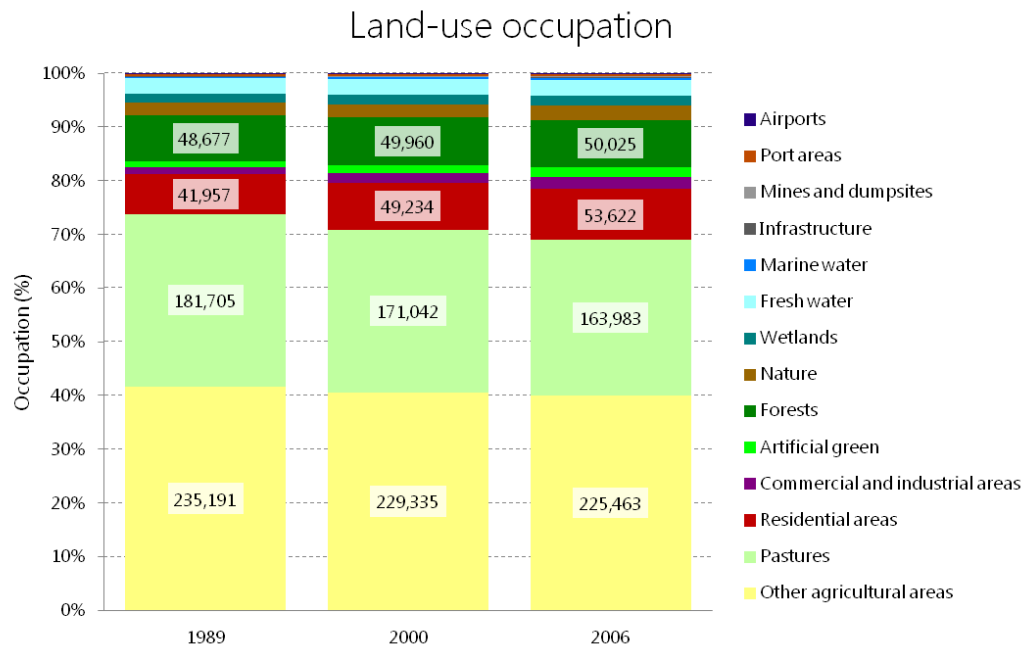


Figure 4.3 – Relative land-use occupation of all land-use types (top) and absolute land-use occupation of urban and natural (bottom). Note for Metronamica users: the latter are modelled as functions in Metronamica.

4.3 Transport network data

A train station map, a highway ramp map and a road network map are available for 2004 (Figure 4.4). The station map consists of points that are divided into two categories: intercity stations and local stations, where intercity stations are defined as stations attended by intercity trains as of 2004. The ramps are also modelled as points and only consist of one category.

The road maps, on the other hand, consist of lines instead of points. City/neighbourhood access roads and larger roads are included. The roads are divided into many segments, ranging from a few meters to a few kilometres. The maps include data about road types, simulated flow levels, simulated congestion levels, simulated speed levels and observed flow levels. Two separate data sources are used to gather this data: the observed flow map is retrieved from Rijkswaterstaat (RWS) and all the other data from Metronamica (where the flow, congestion and speed are simulated). The observed flow data are only available for highways and a handful of other roads. It was decided to use an adapted classification for the roads. This classification is based on the maximum speed.

4.3.1 Road classification

The reason for using an adapted instead of the NWB classification is that the maximum speed is seen as the most generic variable for differentiating between road types (i.e. most likely available in any other road dataset). This is important because the differentiation made here is transferred to Metronamica, which must be able to be used with other applications. The original NWS classification contains quite some variation in maximum speeds in each type of roads (example), making an adapted classification more useful.

The roads were classified into four categories as defined in Table 4.3. The classification is based on the speed limit of the roads, with the exception of level 1. This level is defined as all roads which are classified as 'motorway' in the dataset, regardless of the speed limit (note that speed limits below 100 km/h are present in e.g. sharp turns), because an important defining characteristic of motorways is that they can only be entered at specific points. Furthermore, the motorways create a coherent network throughout the Netherlands, unlike other road types.

When using the simulated traffic characteristics of the roads (i.e. flow, congestion and operating speed), a further classification is used for each of the road levels separately. For example, roads of level 1 are classified into 'low', 'medium' and 'high' flow categories, in 'low', 'medium' and 'high' congestion categories, etc. The intervals for the low, medium and high classes are based on Jenks natural breaks classification method. This method was chosen since the distribution of values tends to be uneven. The Jenks classification method deals with this by choosing the intervals in such a way that the interval borders are chosen at a value where there are few measurements (in mathematical terms, it maximises the variance between the intervals and minimises the variance within the interval at the same time). An overview of the exact intervals can be found in appendix B.

Additional transformations were done as well. In the ramp map, for example, it was seen that some ramps were modelled as two points (one on each side of the road). Since this would disturb the data analysis results, they were merged into one ramp in the centre of the two.

Road category	Road type as classified in NWB	Speed limit in NWB
Level 1	Motorways	-
Level 2	Main roads, access roads (closed declaration), access roads (mixed traffic). private roads city access roads, neighbourhood access roads	(80; 120]
Level 3	Main roads, access roads (closed declaration), access roads (mixed traffic). private roads city access roads, neighbourhood access roads	(50; 80]
Level 4	Main roads, access roads (closed declaration), access roads (mixed traffic). private roads city access roads, neighbourhood access roads	[30; 50]

Table 4.3 – Road classification used in this research

4.3.2 Train station classification

This research uses the same classification as was provided by RIKS. The train stations are divided in two categories: intercity stations and local stations. The first are attended by intercity trains and/or international trains (in 2004), while the second are attended by local trains only (in Dutch terminology: 'sneltreinen' and 'stoptreinen'). Although data on the number of travellers and the number of trains per station were also available, it was decided to use the original classification for two reasons. The first reason is that mainly the *order* of 'small' to 'large' station matters, because (in chapter 5) we are interested in the *trend* that can be seen when going to larger stations. It does, therefore, not really matter how or where the line between categories is drawn exactly. A simple classification would be enough to establish this trend. The second reason is that Metronamica is used for other study areas as well, for which such detailed data are not always available. This research determines whether this relative simple classification would be enough to establish a trend.

4.4 Model application

The Netherlands are divided into 40 COROP regions (used in the regional interaction model) and 413 transport zones (used in the transport model). Please refer to appendix C for maps of the regions and transport zones. Furthermore, the transport network in Metronamica consists of a road network, a waterway network, the locations of highway ramps and the locations of railway stations. These are the same maps as given above, with the exception of the waterway network.

The application includes information in the zoning driver. Natura 2000 data consist of a number of protected natural areas. The model allows no urban developments in this area, while the development of (more) natural land-uses is allowed.

As mentioned above, an adapted road classification is used where the roads are divided in four categories based on the allowed speed (section 4.3).

4.5 Data usage throughout this research

The data above are used in several of the research's steps. An overview of the data use is provided in the Table 4.4. The overrepresentation of land-use developments near transport elements are plotted in graph in the data analysis. The transport and land-use change maps are used in this step. The period of 1989-2000 is used to calibrate the model. An explicit distinction is made between the flow levels of the roads

	Data analysis	Calibration	Validation
Land-use map 1989		✓	
Land-use map 2000		✓	✓
Land-use map 2006			✓
Land-use change map A (1989-2000)	✓		
Land-use change map B (2000-2006)	✓		
Station map	✓	✓	✓
Ramps map	✓	✓	✓
Road type map	✓	✓	✓
Observed road flow	✓		
Simulated road flow	✓	✓	✓
Simulated road congestion	✓	✓	✓
Simulated road operating speed	✓		

Table 4.4 – Data use overview

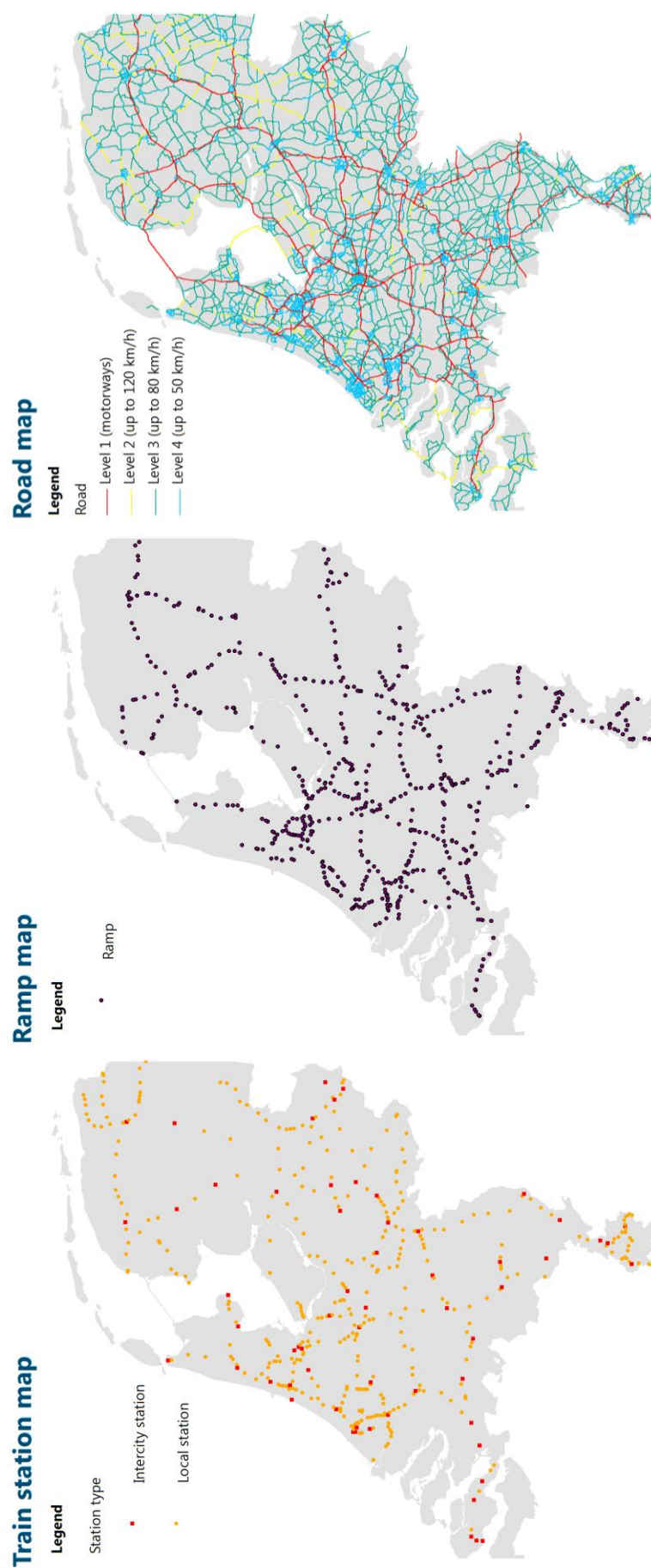


Figure 4.4 – Available transport data for the Netherlands

5 Data analysis

The goal of the data analysis is to examine whether and why land-use developments show an over- or underrepresentation around existing transport elements. The results of the data analysis are used in the next chapter to calibrate the local accessibility in Metronamica.

The attraction (or repulsion) of land-use elements is investigated by analysing the density of land-use developments in an area around the transport element. This density is compared to the density of the case where all land-use developments would be homogeneously distributed over the study area. The results are plotted in graphs which show the overrepresentation to distance. These graphs do not necessarily prove any causal relationship, but shows whether over- or underrepresentation of land-use developments occurs at a certain distance from transport elements.

The methodology leading to overrepresentation graphs is first discussed in section 5.1 . Section 5.2 gives an overview of what overrepresentation graphs were created. The results are presented in section 5.3 and discussed in section 5.3.5.

5.1 Methodology

The local accessibility measure in Metronamica tries to capture the idea that certain land-use types are attracted to transport elements. Bruinsma (1994) noted that the direct distance to transport elements appeared to have equal or greater explanatory power in explaining land-use changes than complex accessibility measure, which is exactly what is captured in the local accessibility measure. Therefore, this chapter focuses on providing an empirical foundation for setting the local accessibility parameters and optimising the measure itself. The goal of this data analysis is therefore to produce overrepresentation graphs, in a similar way as the enrichment factor is used to assess the over- or underrepresentation of land uses in the neighbourhood of a cell by Verburg et al. (2004), see section 3.4.

The first advantage of using a similar method is the possibility to compare attraction to land-use types and transport elements. This makes calibrating the model easier, especially finding a balance between the influence of neighbourhood dynamics and accessibility. Secondly, the resulting graphs can be viewed as a distance decay function for the attracting or repulsing power of a transport element. Thus, the results can be 'translated' to parameter values easily. This fits with the idea that this research does not intend to drastically change the accessibility measure because of limited programming recourses. Thirdly, a tool for analysing overrepresentation in the neighbourhood of cells was already developed by RIKS and adapting it took relatively few resources.

Naus (2009) used a similar method as Verburg et al., but performed the analysis on land-use changes instead of a single land-use map. This accounts for the fact the cells that change over time, are only a small portion of the total number of cells. Looking at the changes also focuses on effects on a relative short term (decades) instead of historical developments (ages). Metronamica is used to simulate changes in comparable time spans. Another difference with the methods mentioned so far, is that we will especially look at densities of land-use developments (i.e. different calculations are used), as is usual in spatial point pattern analysis methods of for example Ripley (Ripley, 1979; Wiegand & Moloney, 2004).

5.1.1 Concept and definitions

The main idea is to check whether the point density in proximity of a point, matches the expected point density in the same area. The concept is illustrated in Figure 5.1, where two fictional situations are presented. A quick visual analysis reveals that in situation A, the points are evenly distributed among the area, while the points are clustered in two sets of four points in situation B. This clustering can be revealed in a mathematical way by observing the area in proximity of each point (in this case a circle, but may be any shape), then counting the number of other points that can be found within this area. In situation A, no points are found

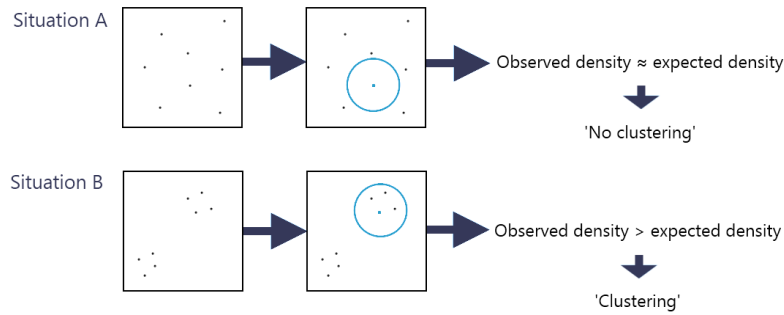


Figure 5.1 – Data analysis concept

within the circle, while in situation B, three points are found. This process is repeated for every point. The expected point density in the circle is determined assuming that all points are homogeneously redistributed over the study area and would thus be same for both situations. Since the observed density is higher than the homogeneous density in situation B, clustering is indicated.

Before using this principle in defining the methodology for this research, some terms related to the method will be defined. They are illustrated in Figure 5.2).

Element refers to any object taken into account in the analysis. They are characterised by distinct geographical locations. In this research, elements represent either land-use cells or transport elements.

The land-use elements are cells in a grid, while the transport elements are either shaped like points (railway stations and ramps) or lines (roads).

Study area refers to the total area of study, while **observation area** refers to an area around an element in which other elements are counted.

We are interested in whether certain transport elements attract land-use types. In other words, the question is whether land-use elements are over- or underrepresented around transport elements. Therefore, two geographic variables are examined at the same time. The procedure of the method can then be described as counting land-use elements in the proximity of transport elements.

The **overrepresentation measure** is the difference between the observed density of land-use elements in the observation area and the homogeneous density assuming a homogeneous distribution of the land-use elements in the study area.

Overrepresentation is present when the observed density is higher than the homogeneous density and **underrepresentation** is present when the observed density is lower than the homogeneous density.

In order to isolate effects at different distances, it is crucial to perform the analysis with multiple observation areas that exclude each other. Figure 5.3 illustrates this principle.

Distance class (c) refers to the number of an observations area, where the area closest to the associated transport element is labelled 1, the second closest in labelled 2, and so on.

Note that the particular shape of the line element observation area is chosen due to the calculation in the tool that is used for calculating the densities, where the determination of whether a land-use element falls into an observation area depends on the straight line distance from this element to the transport area.

5.1.2 Calculation of area of observation area

Before moving on to the detailed description of the overrepresentation measure, the calculation of the area of both the rectangular and circular observation areas will be explained

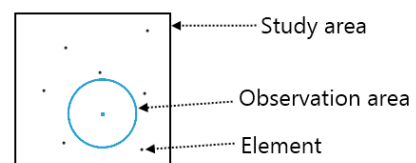


Figure 5.2 – Illustration of definitions

briefly. These calculations will be used in the cluster measure formulas. An observation area is defined by an upper and lower distance limit, see Figure 5.4. These can be defined as:

The **upper limit** ($d_{u,c}$) is the straight line distance from the transport element that defines the outer boundary of observation area c and the **lower limit** ($d_{l,c}$) is the straight line distance from the transport element that defines the inner boundary of observation area c .

For a point transport element, the observation area is a ring of which the area can be calculated by subtracting the areas of the two circles that define the ring:

$$\begin{aligned} A_{oa,ring} &= A_{outer\ circle} - A_{inner\ circle} \\ &= \pi d_{u,c}^2 - \pi d_{l,c}^2 \\ &= \pi(d_{u,c}^2 - d_{l,c}^2) \end{aligned} \quad 5.1$$

For a line element with length l , the observation area can be viewed as composed of two rectangular areas (A) and two half rings (B):

$$\begin{aligned} A_{oa,rectangular} &= 2A_{rectangle} + A_{ring} \\ &= 2l(d_{u,c} - d_{l,c}) + \pi(d_{u,c}^2 - d_{l,c}^2) \end{aligned} \quad 5.2$$

Summarising:

The size of the **observation area** is defined as

$$A_{oa,c} = \begin{cases} \pi(d_{u,c}^2 - d_{l,c}^2), & \text{for transport point elements} \\ 2l(d_{u,c} - d_{l,c}) + \pi(d_{u,c}^2 - d_{l,c}^2), & \text{for transport line elements} \end{cases} \quad 5.3$$

Where:

$A_{oa,c}$	Size of observation area c [cells]
$d_{l,c}$	Shortest distance from transport element to inner boundary of observation area c [cells]
$d_{u,c}$	Shortest distance from transport element to outer boundary of observation area c [cells]
l	Average length of the transport line elements [cells]

5.1.3 Calculation of overrepresentation measure

As explained above, the overrepresentation measure is the difference between the observed density and the homogeneous density and is a function of the chosen observation area distance (c). It was chosen to subtract the two terms in order to be able to identify the absolute differences between land-use types. A division

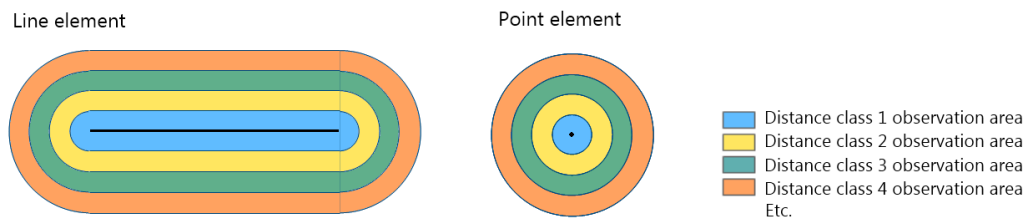


Figure 5.3 – Concept of distance classes

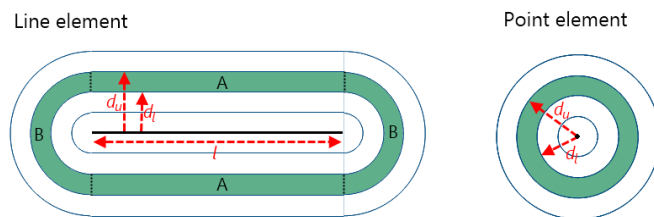


Figure 5.4 – Observation areas are defined by an upper and lower bound

would only lead to an insight in the relative differences. Besides, the measure is independent of the homogeneous density by choosing a subtraction instead of a division. So, in general terms, the following equation can be set up:

$$\Delta\rho(c) = \Delta\rho_{observed}(c) - \Delta\rho_{homogeneous} \quad 5.4$$

Where:

$\Delta\rho(c)$	Overrepresentation measure for distance class c [#cells/area]
$\rho_{observed}(c)$	Observed density in observation area for distance class c [#cells/area]
$\rho_{homogeneous}$	Homogeneous density [#cells/area]

Starting with the right-hand term, the homogeneous density can be described as the total number of land-use elements in the study area divided by the area of the study area. This term expresses the density assuming that the land-use elements are distributed among the area evenly. Similarly, at the left-hand term, the density is expressed by the number of observed land-use elements divided by the area of the sum of the observation areas. More specifically, the observed number of land-use elements is determined by summing the number of observed points over all observation areas. The observation area size is determined in a similar fashion.

$$\Delta\rho(c) = \frac{\sum_{x_c=1}^{n_{tr}} n_{lu,observed,x_c}}{n_{tr}A_{oa,c}} - \frac{n_{lu}}{A_{sa}} \quad 5.5$$

Where:

x_c	Observation area belonging to distance class c
n_{tr}	Number of transport elements in study area [#]
$n_{lu,observed,x_c}$	Observed number of land-use elements within observation area x [#]
$A_{oa,c}$	Total size of observation area c [cells]
n_{lu}	Number of land-use elements in study area [#]
A_{sa}	Total size of study area [cells]

The total size of the observation area is thus calculated by multiplying the size of one observation area with the number of observation areas. However, in reality, observation areas easily overlap each other. Consequently, both the nominator and denominator in the left-hand term give estimated values instead of the actual values. More on the consequences of this approximation can be found in the discussion (section 5.3.5).

Lastly, the results are multiplied by 100 to ease the interpretation of the values. Any positive value of $\Delta\rho$ can be read as 'the number of cells that were found additional to the number of expected cells, per 100 cell area', and vice versa.

The **overrepresentation measure** is defined as

$$\Delta\rho_c = 100 * \frac{\sum_{x=1}^{n_{tr}} n_{lu,observed,x}}{n_{tr}A_{oa,c}} - 100 * \frac{n_{lu}}{A_{sa}} \quad 5.6$$

$$A_{oa,c} = \begin{cases} \pi(d_{u,c}^2 - d_{l,c}^2), & \text{for transport point elements} \\ l(d_{u,c} - d_{l,c}), & \text{for transport line elements} \end{cases} \quad 5.7$$

Where:

$d_{l,c}$	Shortest distance from transport element to inner boundary of observation area c [cells]
$d_{u,c}$	Shortest distance from transport element to outer boundary of observation area c [cells]
l	Length of the transport line element [cells]

The output of the method is therefore one function $\rho(c)$ for each transport element / land-use element combination. These are plotted in graphs of which a typical example is given in Figure 5.5. The attraction of

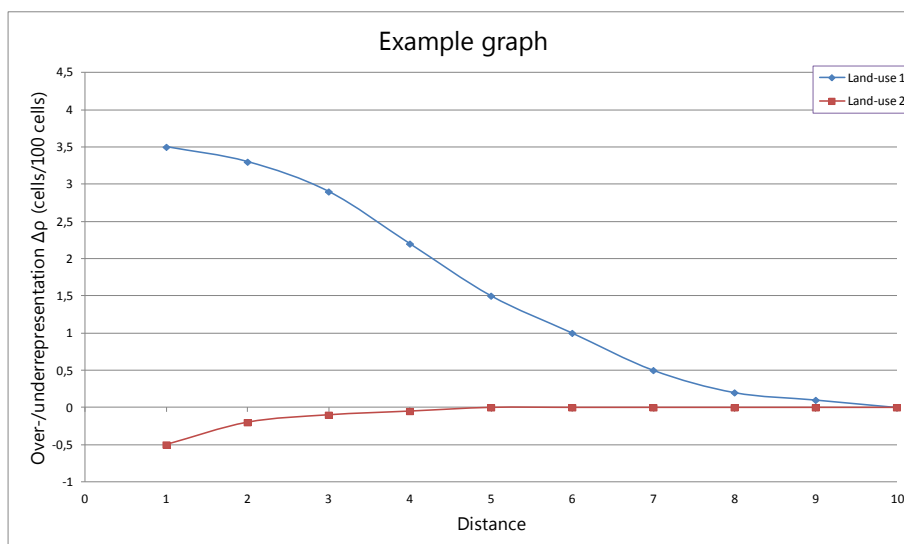


Figure 5.5 – Simple output graph example

all land-use types to a certain transport element is shown in one graph. In the example, land-use type 1 is overrepresented, especially at short distances, while land-use type 2 is slightly underrepresented.

5.1.4 Edge effects

A note should be made to the method described above. In some cases, the observation area will actually fall partly outside the study area. To calculate the correct density, the observation area size should decrease, also known as an edge correction (Wiegand & Moloney, 2004). This would lead to a stronger overrepresentation indication in the results compared to a non-reduced observation area. This is illustrated in Figure 5.6, where it can be understood visually that the overrepresentation in the right-hand situation is stronger. However, edge effects were not included in the calculation for several reasons.

The first reason is the difficulty in efficiently programming the calculation of edge effects in conjunction with the current input formats and the software tool used for calculation the densities. Furthermore, the need for implementing edge effect correction is small, because the study area is quite large (about 200 x 300 km) compared to the observation area radius (maximum 10 km). In other words, there are relatively few instances where an edge correction (Dutch border) would be necessary. Finally, the consequence of ignoring edge effects is a slight underestimation of any overrepresentation values, which is seen as an acceptable error, since it does not lead to any indication of non-existent overrepresentation in the results (the estimation is conservative).

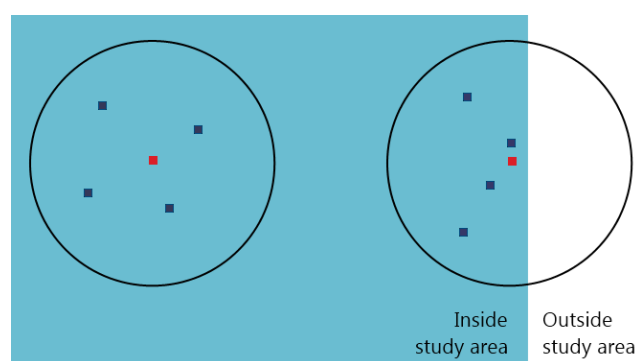


Figure 5.6 – Illustration of the edge effect influence

5.2 Setup

The data analysis is divided in different parts to be able to distinguish between influences from transport element location and quality (Figure 5.7). The first part focuses on the influence of the transport elements itself on land-use changes. This can be seen as a way to test whether the location of the infrastructure has influence on location decisions, by analysing whether overrepresentation is present or not. The second part investigates the influence of the quality of the transport element, by looking at the differences in overrepresentation between categories (that represent different quality levels).

In the first part, the transport elements include railway stations, highway ramps and roads. The question here is whether certain land-use developments are over- or underrepresented along one or more transport element(s). In the second part, it will be investigated whether land-use types are attracted to specific categories within a transport element. The third part entails a specific analysis for the road network, looking at the influence of traffic characteristics (flow, congestion and operating speed). This will give insight into the general influence of traffic on location decision, but also on what traffic characteristic is most important for certain types of roads.

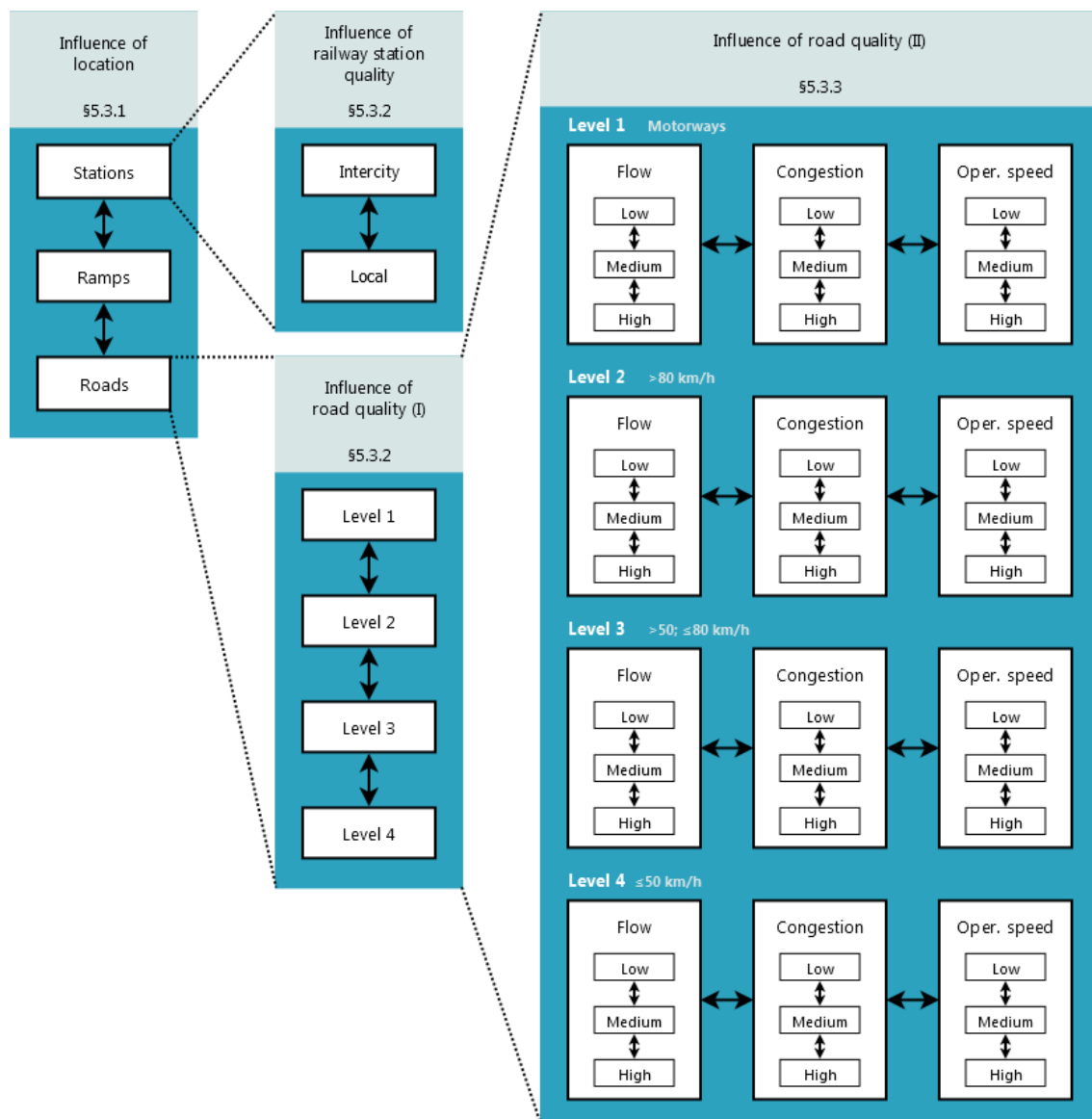


Figure 5.7 – Data analysis overview

Distance class c	Interval $[d_{l,c}, d_{u,c}]$ (cells)	Interval size (cells)	Interval (m)	Interval size (m)
1	[0, 1]	1	[0, 250]	250
2	[1, 2]	1	[250, 500]	250
3	[2, 3]	1	[500, 750]	250
4	[3, 4]	1	[750, 1.000]	250
5	[4, 6]	2	[1.000, 1.500]	500
6	[6, 8]	2	[1.500, 2.000]	500
7	[8, 10]	2	[2.000, 2.500]	500
8	[10, 12]	2	[2.500, 3.000]	500
9	[12, 16]	4	[3.000, 4.000]	1000
10	[16, 20]	4	[4.000, 5.000]	1000
11	[20, 24]	4	[5.000, 6.000]	1000
12	[24, 28]	4	[6.000, 7.000]	1000
13	[28, 32]	4	[7.000, 8.000]	1000
14	[32, 36]	4	[8.000, 9.000]	1000
15	[36, 40]	4	[9.000, 10.000]	1000

Table 5.1 – Distance class overview

The distance classes are defined according to Table 5.1. It is expected that land-use attraction is more sensitive at closer distances from the attraction element. That is why the interval size is set to 1 cell for close observation areas and to 2 or 4 cells for farther observation areas. The maximum distance of 10 km was set after it became clear that the most significant overrepresentation disappears well before this distance.

Additionally, the analysis in Figure 5.7 was repeated on three sub-areas of the Netherlands to check whether the overrepresentation characteristics differ between areas with different land-use patterns. These sub-areas (Figure 5.8) consist of several neighbouring COROP regions and were chosen with expert judgement. They are chosen in such a way that the general land-use patterns are constant within the areas and differentiate as much as possible between the areas. The western area is the most urbanised, the northern area is the most rural and the southern area is somewhere in between. Note that parts of Friesland and Noord-Holland are not included, because the islands belong to the same COROP areas.

5.3 Results

The following is an analysis of the results. Section 5.3.1 contains the results for the infrastructure part. It contains a general discussion about the similarities and differences between transport elements. Then, section 5.3.2 discusses the differences that can be seen when different quality levels within a transport element are studied. The traffic part is discussed in section 5.3.3. Furthermore, we are interested in how the results for ramps compare with those of the motorways (5.3.2.3) and the differences between the time periods for which the land-use change maps were created (period 1989-2000 and 2000-2006, section 5.3.6). The results for both periods will appear to be comparable, which is why the focus is on graphs of period A from here on.

5.3.1 Influence of transport element location

To start the analysis, the different transport elements are compared to each other without subdividing the transport elements into categories. The most important observation is that the urban land-use types (residential, commercial/industrial and artificial green) are

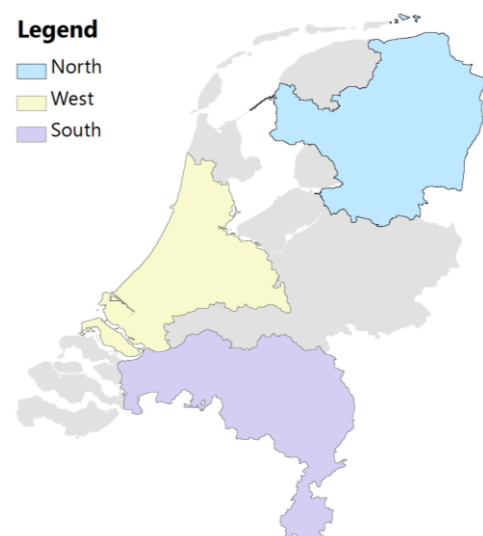


Figure 5.8 – Sub-areas of the Netherlands

overrepresented in the vicinity of all transport elements, as shown in Figure 5.9.

The overrepresentation of urban land-use types close to road transport elements is not a surprise. It is according to theory (e.g. Wegener & Fürst, see chapter 2) and common sense that new residential and commercial/industrial (abbreviated to 'commercial' from here on) activities choose to locate at areas easily accessible by transport means. However, the fact that railway stations seem to attract urban land-uses as well has not been backed up with much empirical evidence yet (Zondag, 2007). The attraction for commercial development to ramps is relatively high at very short distance classes and decreases rapidly, while residential attraction is initially lower, but decreases at a lower rate.

The commercial attraction to ramps can be explained by the preference of companies to be accessible for clients, suppliers, employees, etc. These results are consistent with findings in other research (De Bok, 2007). Similarly, it is also beneficial to be close to a highway ramp for residents, yet being too close to the ramp would involve nuisance. This explains the low value at short distances and the rapid increase directly after.

Furthermore, artificial green is consistently overrepresented over all distance classes. This can be explained by the fact that natural green (e.g. parks) mainly arises within urban areas and is likely correlated with residential and commercial land-use developments. It is therefore expected that the transport elements do not directly attract artificial green, but attract (or are already situated in) urban land-uses, which attract artificial green in turn. Developments in nature and agriculture are consistently underrepresented along all transport elements for the same reason.

5.3.2 Influence of quality: station types and road levels

5.3.2.1 Railway stations

A clear difference can be found when railway stations are categorised into intercity stations and local stations (Figure 5.10). Looking at the intercity stations, it is striking that overrepresentation is relatively small at very short distances (up to 1.5 km). A possible explanation lies in the fact that the area in the direct proximity of large stations is already built-up to full extent and there simply are no more opportunities for new urban cells to emerge. The inertia of urban cells is indeed high, meaning that disappearing is rare. For larger distances, the commercial cells are overrepresented at about 1.5 km and residential at higher distances.

The graphs for local stations are considerably different. Over all distances, residential overrepresentation is by far the highest, even compared to commercial cells. This observation is explainable as well, since local stations serve as the 'home end' of commuter traffic (also keep in mind that the 'local stations' include quite modest village stations).

As mentioned by Zondag (2007), there is not much empirical evidence for a relationship between urban land-use development and public transport elements, except for some specific firm sectors. However, the results of this research do show such a relationship that is at least equally strong as the relationship for other transport elements. Such a relationship is likely dependent on the study area (e.g. these relations may be much smaller in car-oriented regions such as the US) and might therefore be less emphasised throughout literature.

5.3.2.2 Roads

As can be seen in Figure 5.11, there are significant differences between the four road level levels. Commercial attraction is the largest for level 1 (motorways) and is much lower in the other categories, even along other roads with high speeds (level 2). Residential attraction is the greatest for level 4 roads, which are generally around cities (these are the smallest roads available in the data). Additionally, there is a striking attraction for level 1 roads as well, which can be explained in a similar way as the residential attraction to ramps: it is beneficial to be near a highway, but being too close involves too much nuisance.

The 'peaky' shape of the graphs for level 2 roads can be explained by the fact that there are relatively few instances of such roads (1200 segments compared to 5300 for level 1). However, the general shape of the figure indicates important differences between level 1 and level 2 roads and justifies the disaggregation.

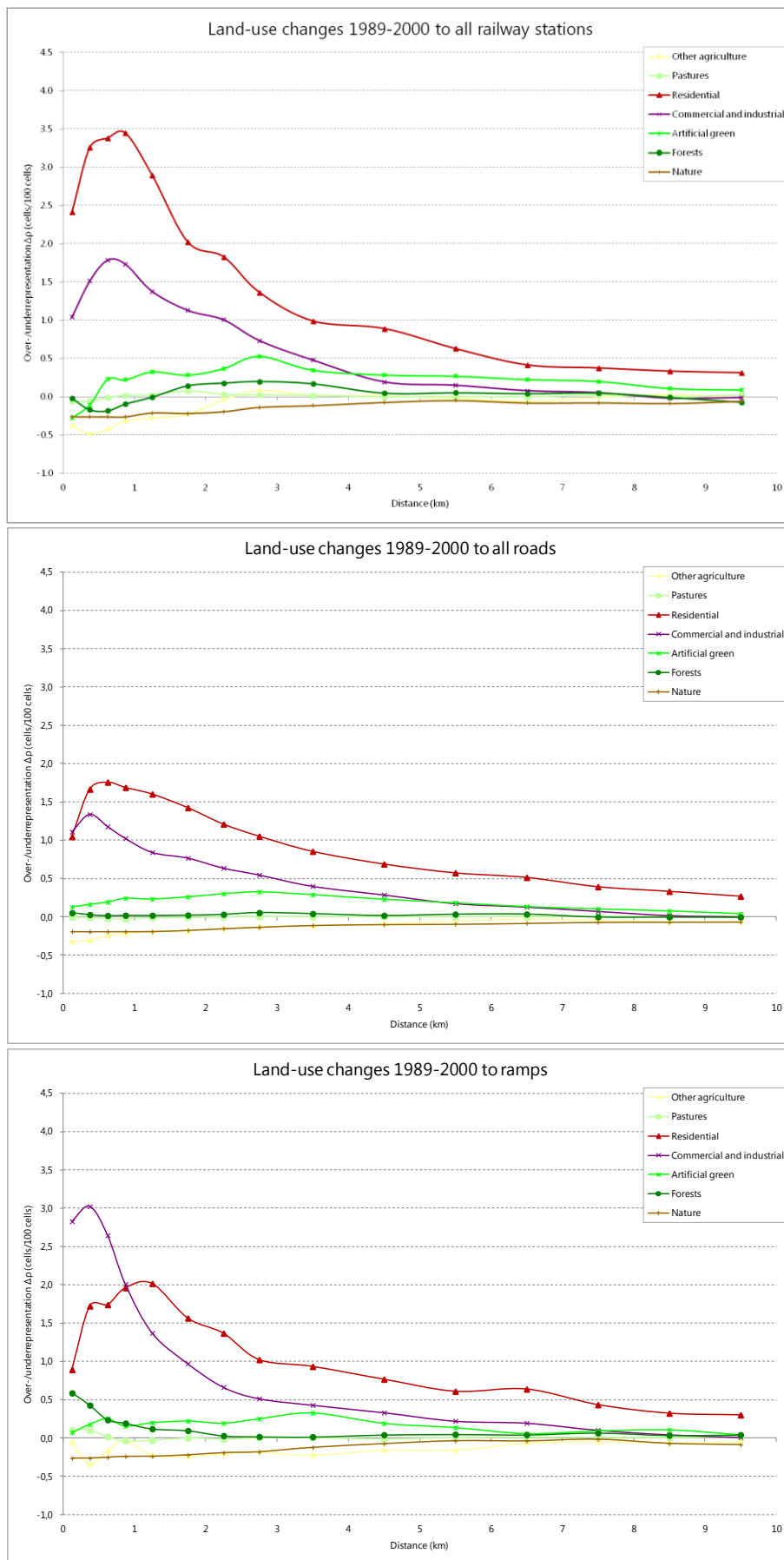


Figure 5.9 – Over- and underrepresentation of land-use changes to stations (top), roads (middle) and ramps (bottom)

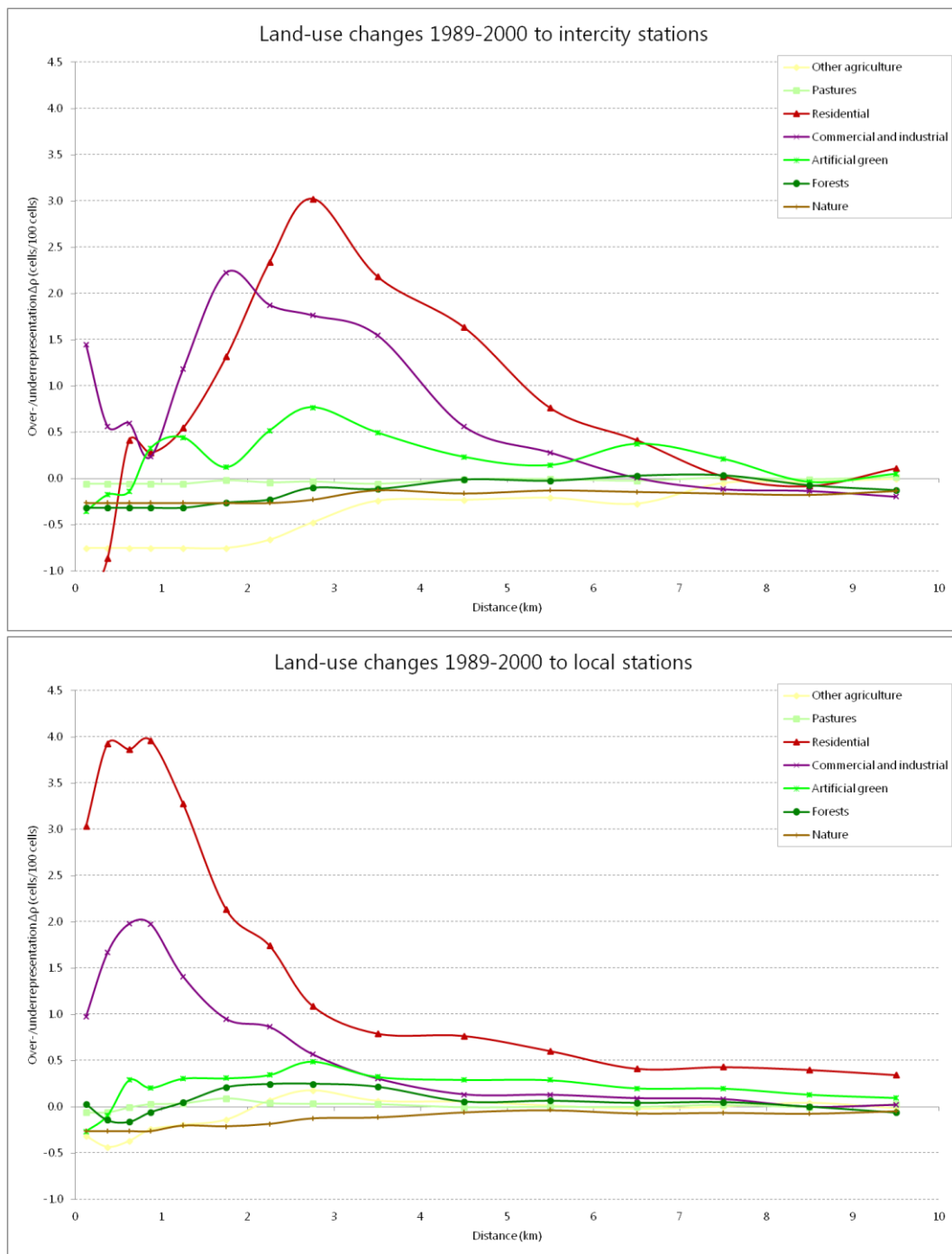


Figure 5.10 - Over- and underrepresentation of land-use change to intercity stations (top) and local stations (bottom)

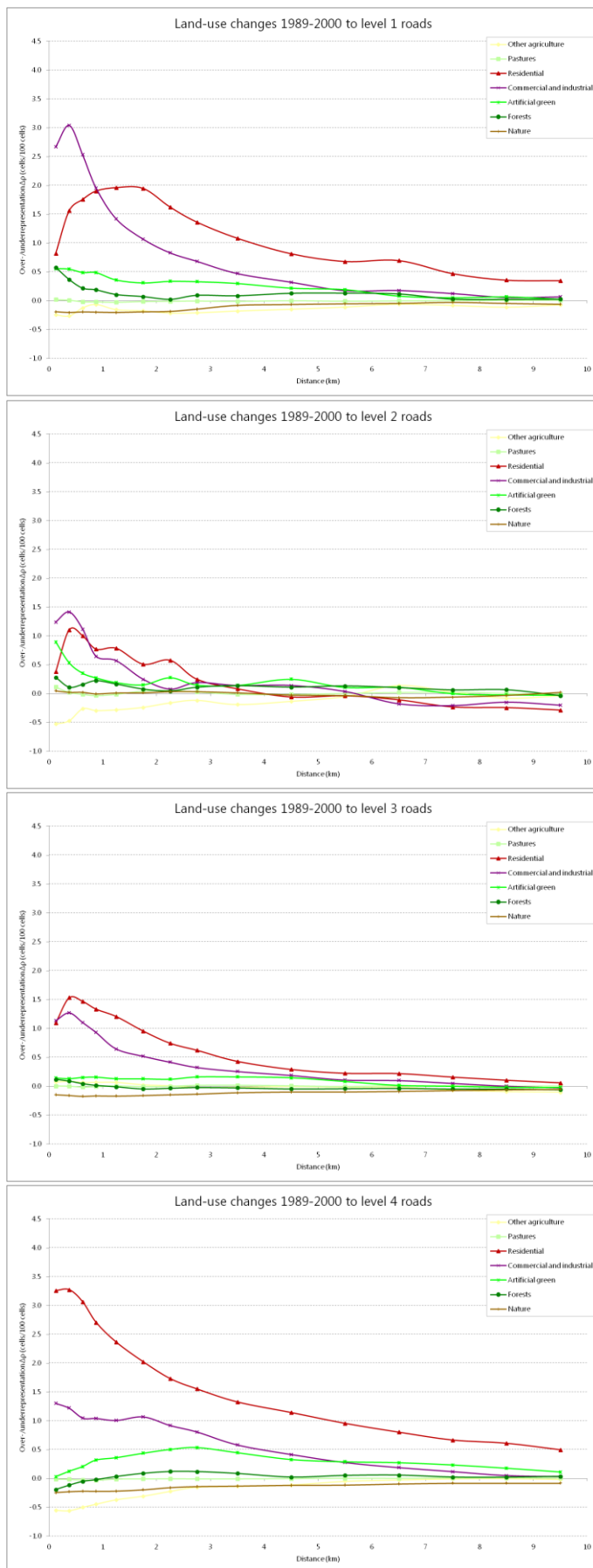


Figure 5.11 - Over- and underrepresentation of land-use change to level 1, level 2, level 3 and level 4 roads

5.3.2.3 Ramps vs. highways

The shape of the ramp graphs is fairly consistent with the highway results (Figure 5.12), which would be natural since ramps are located on highways. In contrast to expectation, ramps do not generate a higher explanatory power than highways. Nonetheless, the implication is that highways are a useful and reliable substitute for modelling land-use change around ramps. Using ramps in modelling practice has the benefit of using point elements instead of line elements (see section 5.5).

5.3.3 Influence of quality: traffic characteristics

The influence of traffic characteristics is investigated per road level (instead of for the whole network), since road level and traffic characteristics are correlated to a high degree, while we are interested in the separate influence of traffic characteristics. Besides, it appeared that the influence of traffic characteristics contains important differences between road levels.

Although they are analysed separately, the road characteristics are not independent of each other. The flow is actively modelled in Metronamica using the classic four-step model (Ortúzar & Willumsen, 2001). The congestion follows as the flow/capacity ratio and the operating speed is finally derived from the congestion (free flow capacity is reduced depending on the amount of congestion). The results are therefore also related to a certain extent. However, one of the fields of interest is to determine which of these road characteristics actually entails the greatest power in explaining land-use changes.

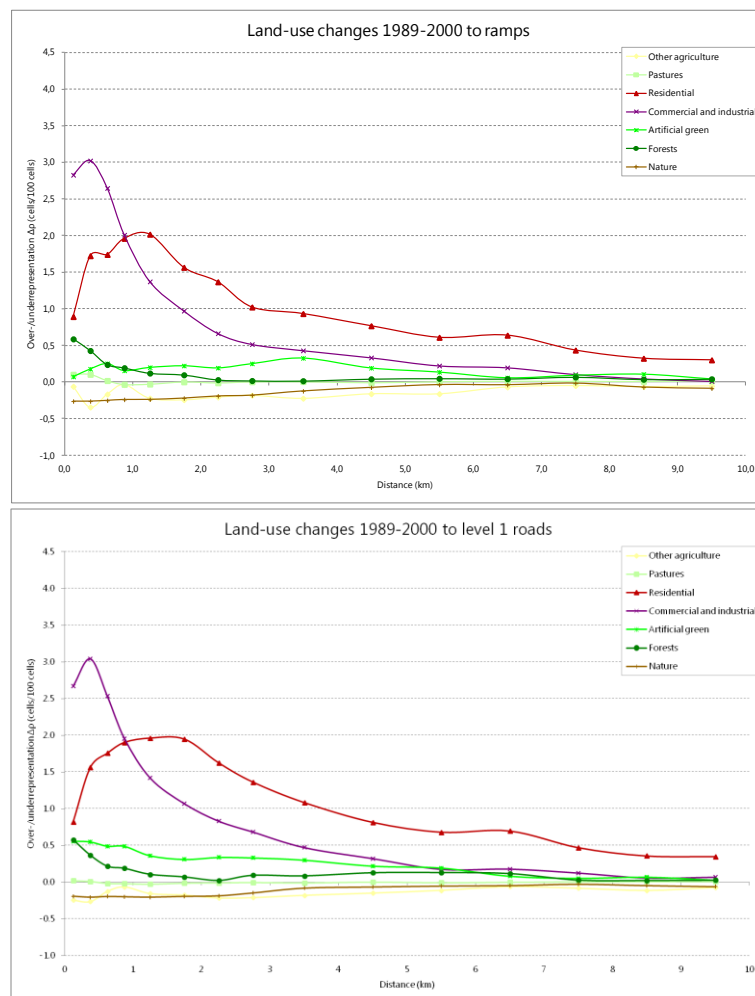


Figure 5.12 – Over- and underrepresentation of land-use changes for ramps (top) and highways (bottom)

The operating speed results were found to be extremely similar to the congestion results, so the analysis below is focused on congestion and flow. The similarity between congestion and operating speed can simply be explained by the fact that one is directly calculated out of the other.

5.3.3.1 Level 1 roads (motorways)

Both residential and commercial land-use developments show a relatively high overrepresentation along all level 1 roads, see Figure 5.14. Commercial land-use generally favours low flow motorways, while residential developments prefer high flow motorways. When looking at congestion, the same patterns can be seen with the exception that residential developments do not prefer high congestion roads.

These results can be explained by looking at the map for residential and commercial developments Figure 5.13. The residential developments tend to cluster together and arise mainly in the Randstad, where the highest flow roads are located as well. In contrast, commercial developments seem to appear throughout the whole study area. This indicates that residential actors might be stronger influenced by the location of current residential land than by the flow or congestion of the road. However, the overrepresentation of residential developments along highly congested roads is significantly lower than along high flow roads, leading to the conclusion that residential developments at least avoid congested roads to some extent.

5.3.3.2 Level 2 roads (> 80 km/h)

The graphs for level 2 (Figure 5.15) are quite uneven, which could be explained by the fact that there are too few roads in this level for the results to be more consistent. Nonetheless, all overrepresentation is lower than level 1 roads, indicating that the preference for highways is higher than other high speed roads, regardless of the traffic characteristics. Another major difference to level 1 roads is that commercial cells are more overrepresented along high flow roads compared to lower flows. This might be explained by the fact that since we are now looking at smaller roads, high flows are less sensitive to congestion. The congestion levels of the roads do not seem to influence the preferences of both residential and commercial land-use changes.

Land-use changes 1989-2000

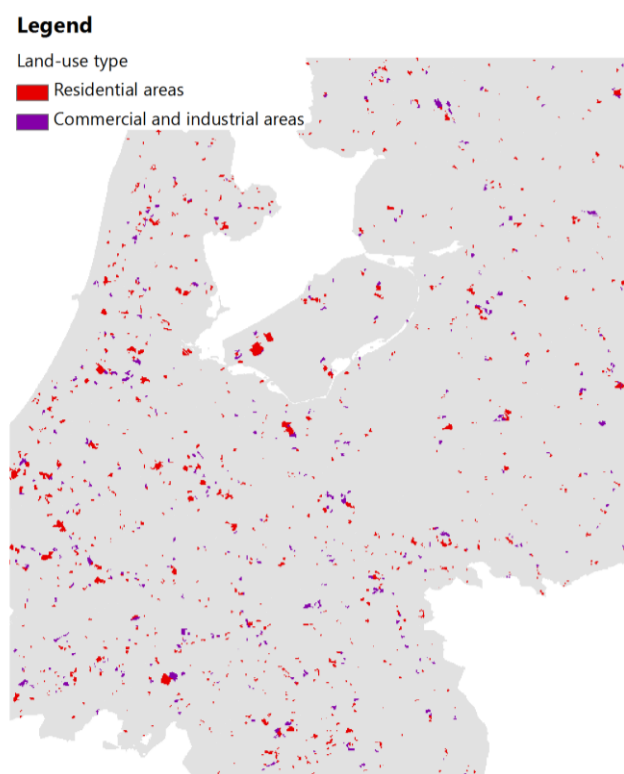


Figure 5.13 – Residential and commercial/industrial land-use changes between 1989 and 2000

5.3.3.3 Level 3 roads ($> 50; \leq 80$ km/h)

The road flow seems to be the factor with the most explaining power for choosing one level 3 road over the other (Figure 5.16). The higher the flow, the more likely it is that urban land-use developments will appear. A reason for this is that these roads are regional roads where there is still room for development. Since there are (in general) less congestion problems than on e.g. highways, the urban land-uses choose for the high flow roads. Similarly, high congestion roads show a higher overrepresentation of urban developments, although this difference is not nearly as large as the difference between the flow levels. This observation would lead to the conclusion that higher flows are preferred except when these higher flows lead to congestion.

5.3.3.4 Level 4 roads (≤ 50 km/h)

The graphs for level 4 roads appear to have odd shapes at first sight (Figure 5.17). The commercial developments do not have an outspoken preference for a specific flow or congestion level. The graphs for new residential developments, however, are not that straightforward. Low flow roads are preferred at shorter distances, while high flow roads are preferred for distances longer than 2 km.

A possible explanation is that living very close to a busy (i.e. high flow) city road has several negative aspects, such as noise and air pollution. Living slightly further away negates the disadvantages while maintaining the benefits of living close to a high flow road. However, it is still not explained why the high flow roads become the most attractive only after 2 km (the disadvantages would disappear after a few hundred meter). An additional factor that might play a role is that high flow roads are generally ring roads around the city, while low flow roads penetrate deeper into the city. Note that the latter might be an equally strong or stronger variable for explaining differentiation between level 4 roads.

5.3.4 Sub-areas

The analysis was repeated for three sub-areas of the Netherlands (see Figure 5.8): one relatively urban (west), one relatively rural (north) and one in between (south). It must be noted that this is only a rough analysis with the goal to check whether differences can be seen between the areas, not to analyse these differences and use them. A more detailed and thorough analysis should be performed to gain more insight in the differences between sub-areas. This is not performed here, because these differences cannot be implemented in Metronamica yet (this could, however, potentially improve the simulation results).

Indeed, the resulting graphs show some notable differences. An example is given in Figure 5.18, where the overrepresentation graphs for ramps are shown for all three sub-areas. The most striking observation is that commercial overrepresentation is by far the highest in the north and the lowest in the west. This might be explained by the fact that in the north, there is still more space available. Consequently, actors have more freedom to choose locations, which would be closer to transport elements.

It would be false to conclude that the overrepresentation in the north is higher because the transport network itself is less dense. Note from formula 5.6 that this is taken into account in the calculation, because the observation area size $A_{oa,c}$ is smaller for less dense networks.

Since the sub-areas are smaller, the number of available transport elements within these areas can be small. This is, for example, the case with intercity stations. The resulting graphs are extremely spiky, thus making a descent analysis impossible.

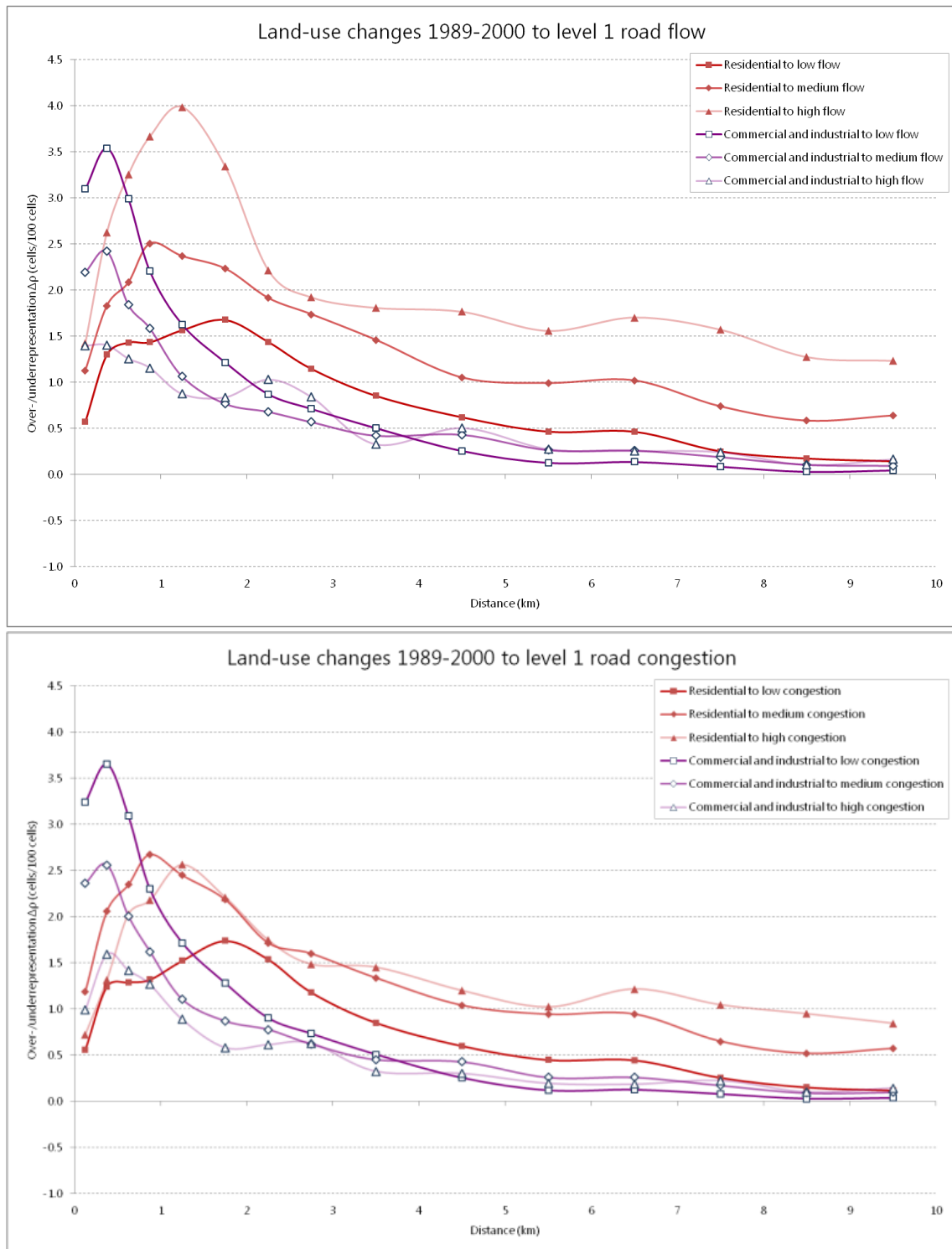


Figure 5.14 – Over- and underrepresentation of residential (red) and commercial (purple) land-use changes to level 1 roads, for different flow levels (top) and congestion levels (bottom). Lower flow and congestion levels are indicated by darker colours.

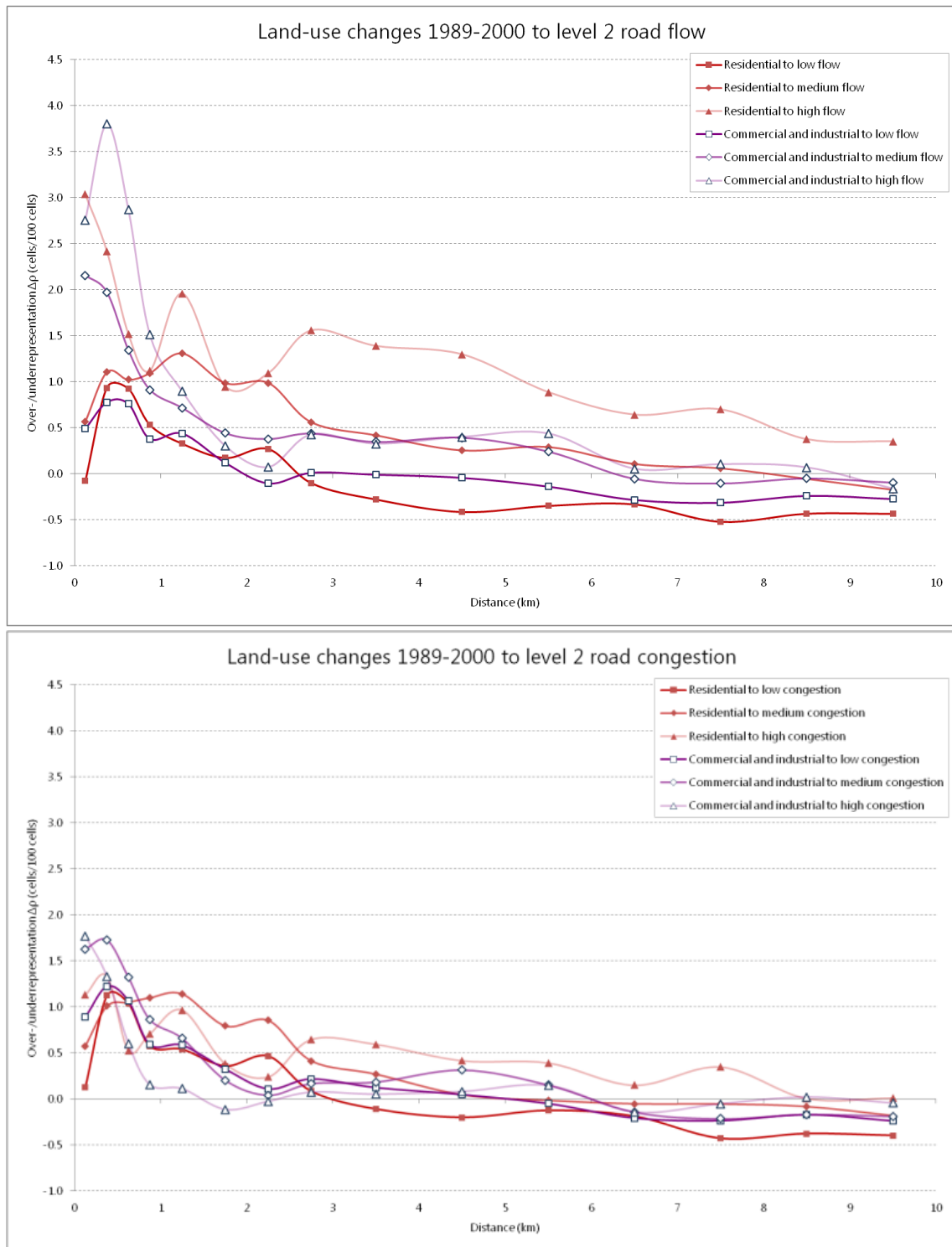


Figure 5.15 - Over- and underrepresentation of residential (red) and commercial (purple) land-use changes to level 2 roads, for different flow levels (top) and congestion levels (bottom). Lower flow and congestion levels are indicated by darker colours.

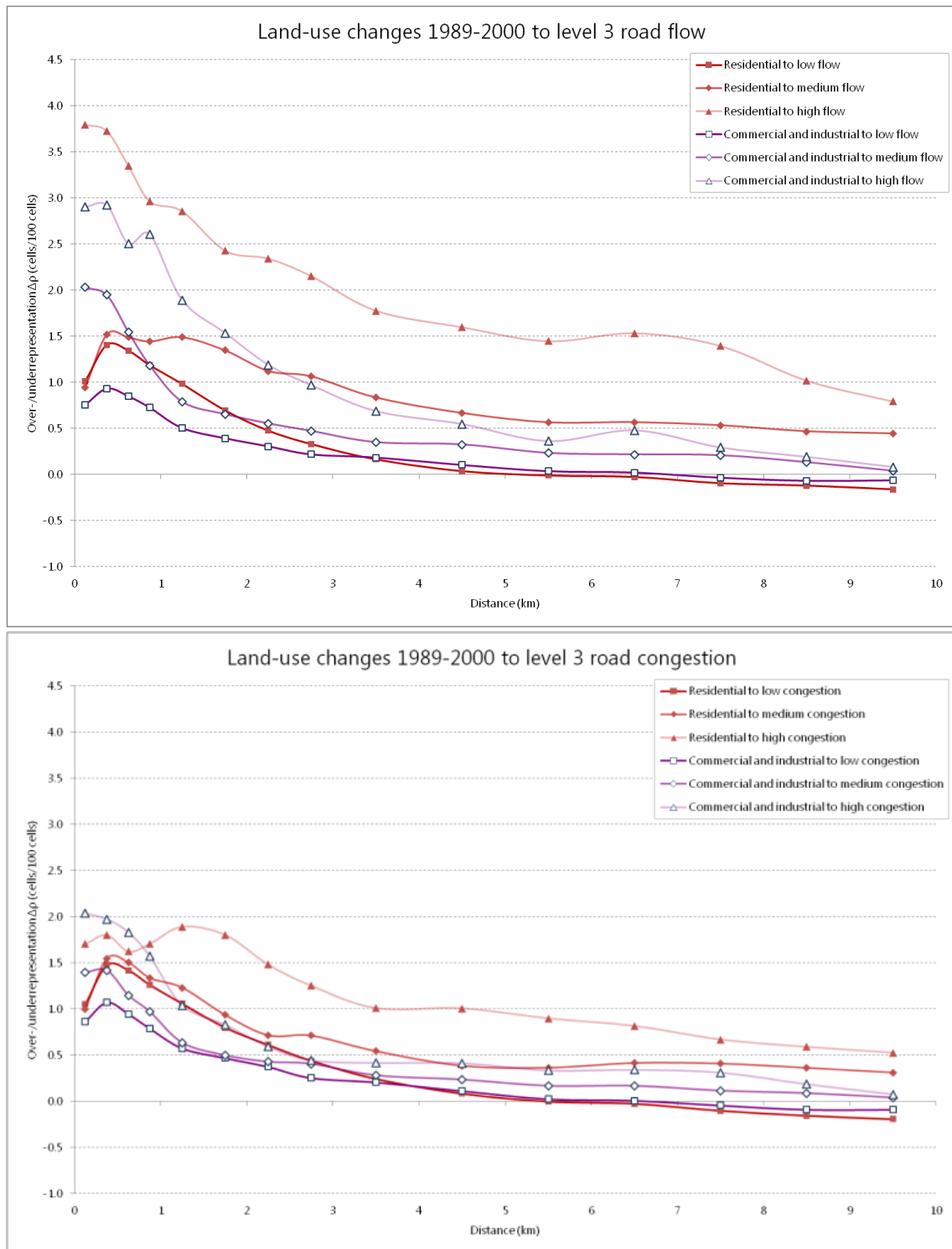


Figure 5.16 - Over- and underrepresentation of residential (red) and commercial (purple) land-use changes to level 3 roads, for different flow levels (top) and congestion levels (bottom). Lower flow and congestion levels are indicated by darker colours.

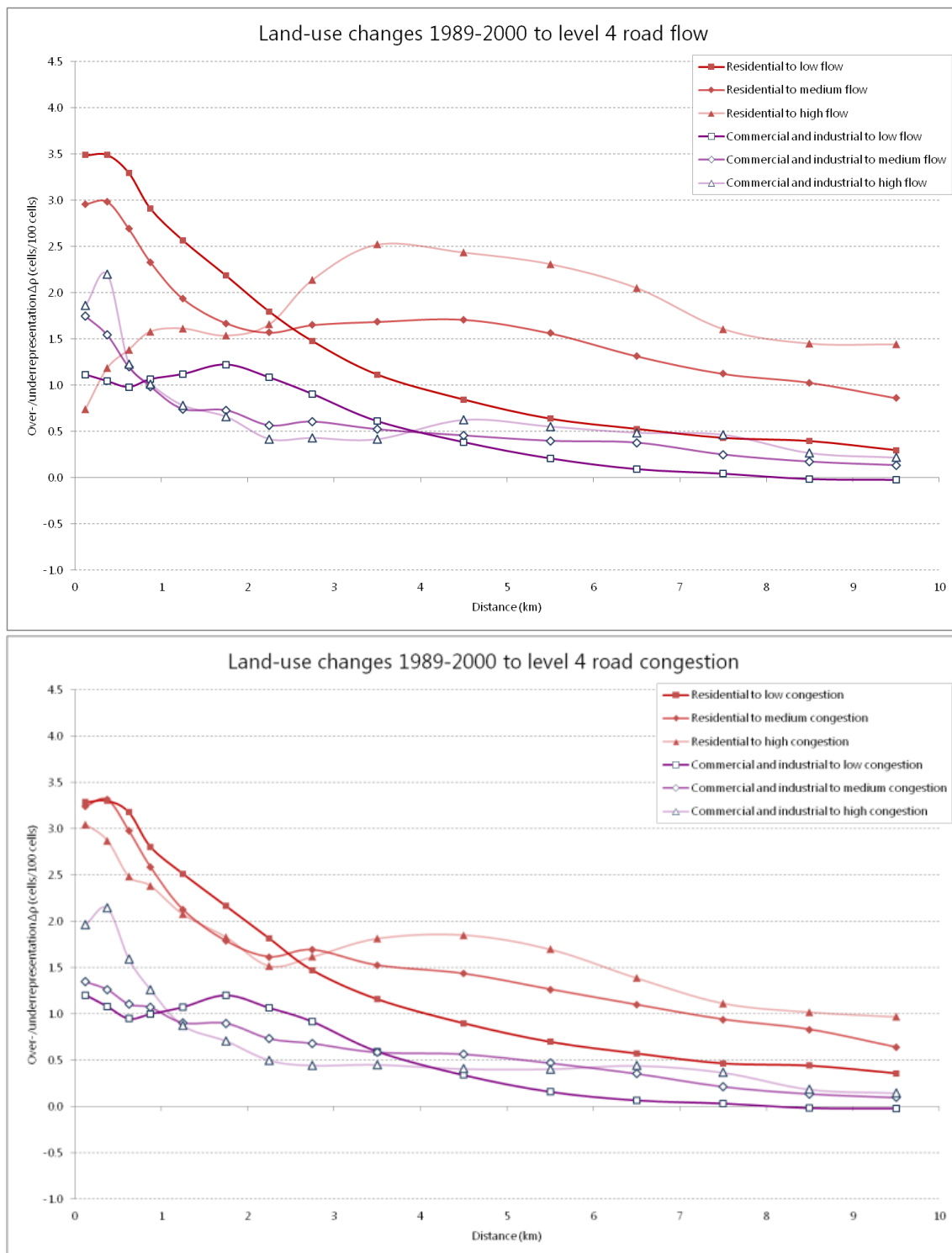


Figure 5.17 - Over- and underrepresentation of residential (red) and commercial (purple) land-use changes to level 4 roads, for different flow levels (top) and congestion levels (bottom). Lower flow and congestion levels are indicated by darker colours.

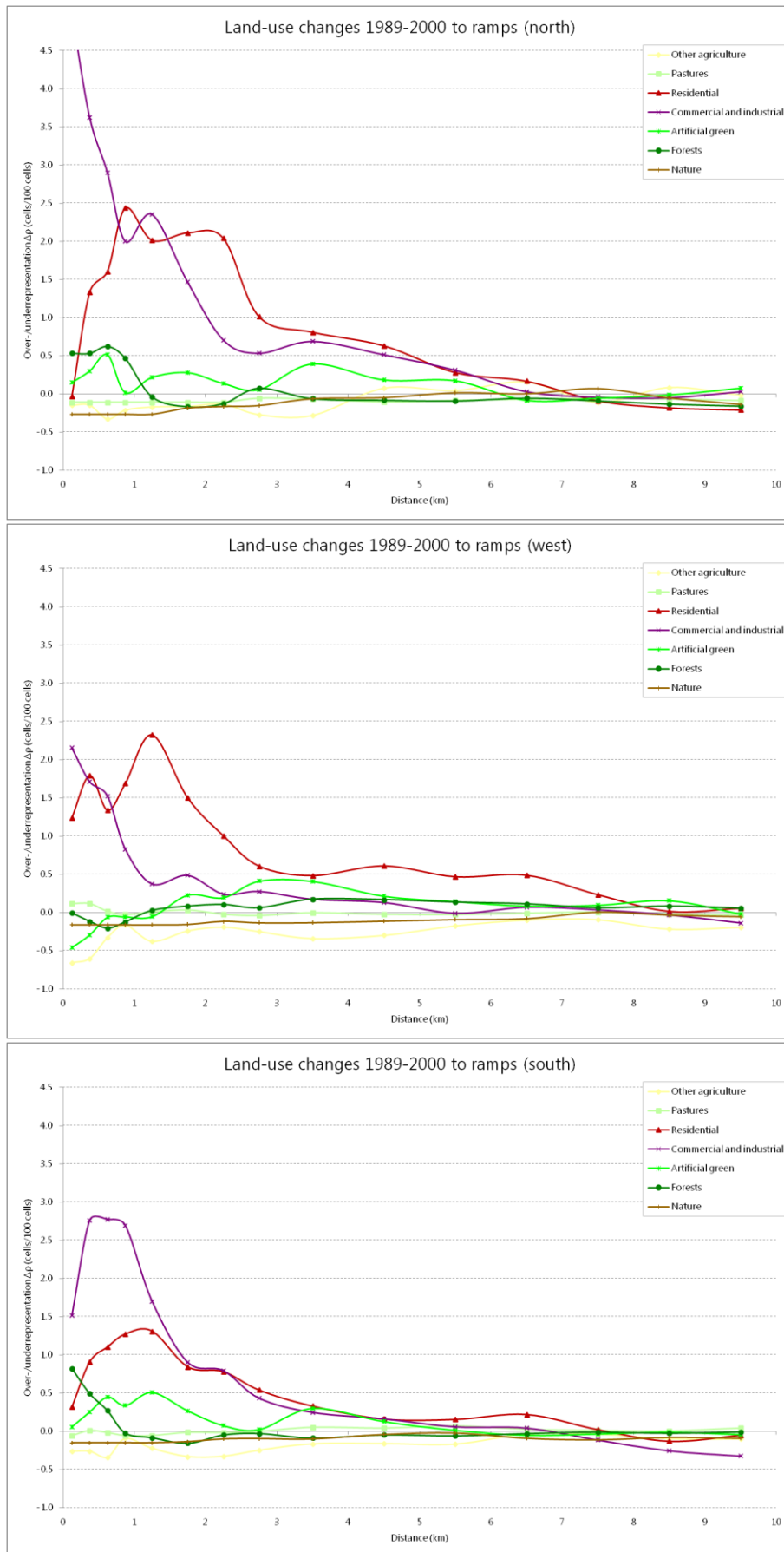


Figure 5.18 – Overrepresentation to ramps for three sub-areas of the Netherlands

5.3.5 Observed road flows

Overrepresentation graphs were also created for observed flows. The main function of this analysis is to check whether the simulated flows graphs are a decent indicator for the actual flows. The observed flows are only known for motorways and some other roads. Comparing the graphs for observed flows to simulated flows, similar patterns are discovered. The most important conclusion is that the output of the transport model is realistic and so are the overrepresentation graphs that rely on it.

5.3.6 Differences period A and B

Despite some minor differences, the results of period A (1989-2000) are generally consistent with those of period B (2000-2006). The most noticeable difference is that the overrepresented land-uses (i.e. residential and commercial) contain more extreme values in period A. This difference might be explained in two ways: firstly, period A entails a larger time period, thus there are more instances of the attracting effect to be found. Secondly, the trend that residential and commercial land-uses are attracted by transport elements might simply have become weaker over time. However, since the differences are minor and no further information is available, there is no reason to assume that (part of) the data analysis results do not apply to more recent periods.

5.4 Discussion

5.4.1 Overrepresentation and causality

It would be easy but erroneous to assume that land-use developments are caused by the proximity of certain transport elements, solely because overrepresentation is shown. Although the hypothesis of the data analysis assumes the existence of such a relationship to some extent, other factors play undoubtedly a role as well. These might include zoning plans that are located next to a certain transport element and preferences to be located near other non-transport elements. An expert evaluation of the results with specific attention for causality is therefore of great importance and was provided in the discussion of the results as much as possible.

An example where overrepresentation is shown and direct causality is likely, are commercial developments near motorways. For example, zoning plans for business parks might allocate new commercial developments next to the motorway. The question is whether this allocation is caused by the zoning plan or by accessibility to the motorway. The latter probably played a role in deciding on the location of the business park in the zoning plan, so, in this case, the reason for the overrepresentation would be a combination of zoning and proximity to motorways.

An example where overrepresentation is shown, but where direct causality is questionable, is the overrepresentation of artificial nature (to all transport elements). As mentioned before, it is more likely that artificial nature develops near (or in) urban areas, because their function is mainly to provide some green within urbanised areas. Consequently, artificial green developments are seen near transport elements, because urban developments took place in these areas earlier.

The challenge lies ultimately in how to model each of these effects, of which the one is more direct than the other. In the first example, it is reasonable to assume that commercial cells would develop in areas accessible by motorways even without the zoning plan(s). The inclusion of commercial attraction to motorways as a direct effect in the model would thus be acceptable. The challenge is more difficult in the second example. Since the relationship is indirect, it should not be modelled ideally. However, since overrepresentation is present, the simulated results might actually improve in practice. In the case of Metronamica, artificial green's attraction to urban cells can be modelled directly in the neighbourhood driver, so there would be no need to include the overrepresentation in the accessibility driver.

Some tests were done where land-use developments designated by policy were omitted from the analysis. More specifically, these zoning plans included the Vinex locations for residential developments and a variety

of provincial and municipal plans from the 1990s which were available in 'De Nieuwe Kaart' (Nirov, 2010). All these plans together covered a small part of the total developments (about 15% of residential and 3% of commercial cells; see Appendix F for map). The overrepresentation graphs showed a slight tendency to move towards the x-axis as a result. Noting that the set of policy plans probably does not cover the whole extent which was relevant for the 1990s, it is again expected that accessibility also played a role in the formulation of policies.

5.4.2 Translation of the empirical results to model parameters

Similar to the neighbourhood rules, the overrepresentation graphs can be used to calibrate the local accessibility. The distance decay parameter is indeed roughly similar in function and shape to the decay in overrepresentation. Some considerations were discussed by Verburg et al. (2004) and are applicable for the translation of overrepresentation graphs in this chapter to distance decay parameters in Metronamica as well (which will be done in the next chapter).

First of all, as stated several times by now, accessibility is not the only explaining factor for land-use determinants. Since not all overrepresentation necessarily contains a causal relationship, it may not be wise to directly translate the result graphs into distance decay parameters. Besides, the graphs are based on changes in an eleven-year time span while the distance decay parameters are used to model yearly changes. This introduces another reason why the distance decay functions in Metronamica can look differently from the empirical overrepresentation graphs.

Nonetheless, the overrepresentation graphs offer a better starting point for the calibration of the local accessibility distance decay parameters than a calibration that solely relies on expert judgement. Verburg et al. proposed a conceptual framework which explains this process (Figure 5.19). This framework can be adapted and applied to the calibration of this research as well.

5.5 Conclusions

The empirical study in this chapter shows that attraction from transport elements to urban land-use developments is likely. At least, an overrepresentation in development density is shown around these elements. However, overrepresentation does not necessarily mean attraction, since an overrepresentation does not indicate a direct causal relationship in all cases. Nonetheless, after expert judgement of the

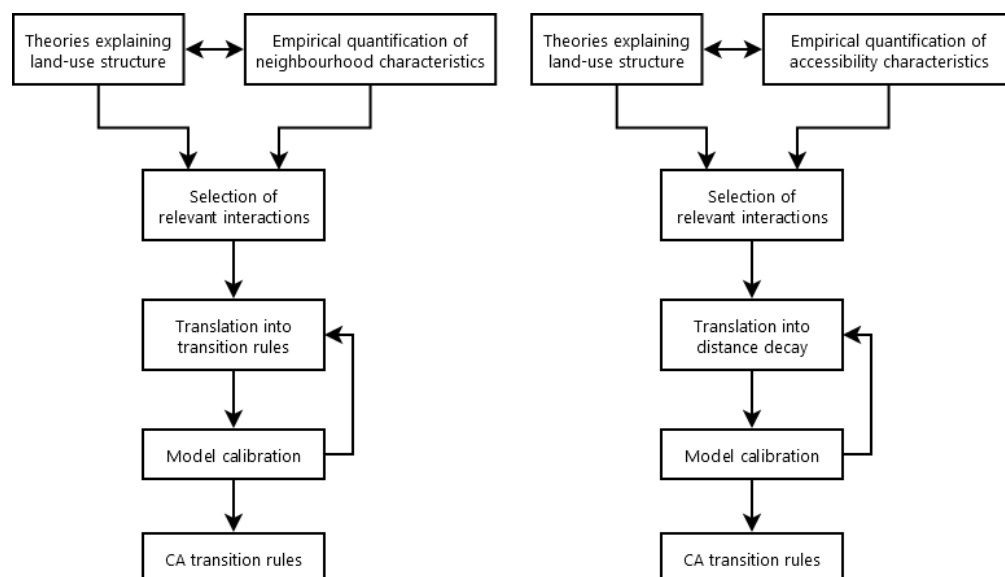


Figure 5.19 – Translation of empirical results to model parameters. Left: the original concept for neighbourhood dynamics (Verburg, de Nijs, Ritsema van Eck, Visser, & de Jong, 2004). Right: analogous concept for accessibility.

causality, the results can be used as a basis for the calibration of the local accessibility measure of Metronamica.

On a general level, it was seen that residential and commercial land-use developments are attracted to all of the investigated transport elements. The extent of the attraction differed between these elements, but also within these elements.

The differentiation between main and local stations indicates that larger stations attract more commercial development. However, this development does not take place directly next to the stations, but starts at about 1 km. Changes do not appear at short distances, because the areas around intercity stations are already well-developed urban areas in 1989. Changes will therefore not appear at very close distances. This does not mean, however, that there would be no attraction in the (somewhat unrealistic) case of a intercity station surrounded by non-urban land-uses. Attraction at close distances should therefore be modelled in Metronamica. Additionally, different parameters should be set between main and local stations.

Differences were also seen between road levels, where commercial attraction becomes smaller as the level decreases. Residential developments are mainly attracted by low level roads and, surprisingly, motorways. However, it is more interesting to analyse the differences between traffic characteristics. For all road levels, flow seems to be the variable with the highest explaining power. However, there are important differences between the road levels. For example, low flow highways are more popular for commercial land-uses than high flow motorways. High flow roads are more sensitive to congestion, which might explain the reduced attraction for commercial developments. The opposite effect can be seen for lower level roads: mainly the high flow roads attract commercial developments. Ultimately, it is expected that setting different parameters for different flow levels will generate a better simulation of commercial land-use patterns.

All in all, the findings in this chapter form an empirical basis to set the local accessibility parameters in Metronamica. The differences between flow and congestion levels indicate that a better simulation might be achieved when these differences are included in Metronamica. The next chapter uses these ideas to check whether the simulation can indeed be improved.

5.6 Reflection

5.6.1 Differences between study areas

The influences of land-use determinants can differ greatly between regions (Bürgi & Turner, 2002), which is no different for the influence of transport systems. Due to cultural, social, economic and other differences, the influence of a certain transport element might be different in one area compared to the other. The influence of stations, for example, is expected to be significantly less in more car-oriented societies.

Another difference between study areas is the relative importance of accessibility in the total potential compared to the other three drivers (suitability, zoning and neighbourhood effects). In areas with a well-developed transport network and a relatively consistent level of accessibility over the area, the accessibility plays a smaller role than in areas where the transport system is less developed and accessibility levels differ greatly in the area (e.g. Germany vs. Spain). This is an effect that has been observed by RIKS in applying and calibrating Metronamica for many different areas.

As a final remark, research by Verburg et al. (2004) has shown that explaining variables for land-use patterns can differ in magnitude between regions of the Netherlands. Further insights in attraction to transport elements might be gained when the analysis is repeated after dividing the Netherlands in smaller areas.

5.6.2 Line elements and overlapping observation areas

One methodological issue with formula 5.6 lies in the combination of transport line elements and the observation area shapes. Considering a road segment end, it is always connected to one or more other lines. Consequently, the observation areas overlap, which causes all points in this overlapping area to be counted

multiple times (hence the factor $\sum_{x_c=1}^{n_{tr}} n_{lu,observed,x_c}$). The reason for this formulation is that the software tools used do not allow for the use of the actual values without significant (and undesired) alteration of the algorithms. This is discussed in detail in appendix G, but it mainly becomes a consideration for line elements, since many overlap is present.

In short, this error would lead to an overestimation of the overrepresentation only if the area where multiple observation areas overlap, is more attractive than the area with one observation area. This overestimation increases as the number of overlapping elements increases. However, this issue does not seem to cause a problem in practice for two reasons. Firstly, the overrepresentation decreases for larger distances, while the error would lead to overestimating the values for the largest observation areas. Secondly, the results for ramps and motorways are extremely similar, while one is a point element and the other a line element.

5.6.3 Aggregation of commercial and industrial sectors

The land-use maps only have one land-use type for commercial and industrial sectors. These include shops, business parks, factories, etc. In reality, the location decisions made for these land-use types are made by different stakeholders and the behaviour is therefore expected to differ. Researchers have indicated that preferences to be located near motorways differ greatly among commercial and industrial sectors (De Bok, 2007; De Bok & Sanders, 2005; Leitham, McQuaid, & Nelson, 2000). Ideally, these different sectors should have been included in the analysis.

Whether a disaggregation between industry sectors is desirable, depends on the application of the model. For example, it might be beneficial for study areas that are mainly urban and that are modelled on a small scale. In those cases, it might be chosen to respect the differences in attraction preferences in the calibration of the model and, thus, not to aggregate the data, provided the available data and recourses allow such a calibration.

The Netherlands application used in this research, though, is on a large scale and contains urban, rural and natural areas. Combined with the fact that the commercial land-use type consists of relatively few cells, patterns would be more difficult to recognise when further disaggregation would be used.

5.6.4 Analysing infrastructure changes

It would be interesting to perform a similar analysis, but focussing only on *changes* in the infrastructure network in the 1989-2000 period instead of the whole network. This might add additional information to the results discussed above. Unfortunately, the older network data that were available was drawn less precise than the 2004 data that was used to perform the analysis. Comparing the older and the 2004 maps to identify the changes, would therefore be too time-consuming.

6 Local accessibility calibration and validation

This chapter deals with the question how the empirical results of the data analysis (chapter 5) can be used in Metronamica. Three simulations are set-up for the period 1989-2000: one without the influence of accessibility, one with the influence of accessibility and one with differentiation to transport element quality. This is described in section 6.1. For validation purposes, the simulations are repeated for periods 2000-2006 and 2006-2030.

Indicators in three categories (cell-by-cell comparison, patterns comparison and visual comparison) are used to assess how well these simulations fit the actual land-use map of 2000, see section 6.2. The results are shown and discussed in section 6.3.

6.1 Simulation set-up

Three simulations (Table 6.1) are carried out and compared together with a set of indicators that will be described below. In *simulation I*, the accessibility is turned off in the land-use simulation. This means that the transition potential formula is the following instead of formula 3.1:

$$TP_{f,c} = \begin{cases} S_{f,c} * Z_{f,c} * N_{f,c} * R & \text{if } N_{f,c} \geq 0 \\ (2 - S_{f,c} * Z_{f,c}) * N_{f,c} * R & \text{otherwise} \end{cases} \quad 6.1$$

The goal of running a simulation without accessibility is twofold. Firstly, it can be investigated what the effect of accessibility is on the land-use simulation. Secondly, it can be seen how overrepresentation graphs look like when modelling without accessibility. This will tell us something about what effects are already modelled through other drivers and where accessibility should correct the overrepresentation patterns.

Accessibility is included in *simulation II*. No differentiation between the quality of transport elements is included yet. In *simulation III*, different weights and distance decay parameters are set for different quality levels of train stations and roads. The goal of the second simulation is to see the influence of accessibility on land-use simulation and the goal of the third simulation is to investigate what the effects are of setting different parameter values for different quality levels.

The three simulations were carried out *ceteris paribus*, meaning that only the parameters in the local accessibility component were changed (the distance decay and relative weights, see section 3.5) and that all others were kept equal. Appendix H gives a short overview of the most important other parameter settings.

The first and second simulations were already calibrated by RIKS and were not adjusted. The third simulation introduces many new parameters and was subject to a more comprehensive calibration. This procedure is described in section 6.2 and the results in section 6.3.

6.1.1 Definition of transport element quality

The quality of transport element is defined as follows (Table 6.2). For train stations, the same two classes are used as before in the data analysis: the intercity stations (attended by intercity trains in 2004) and local

	Simulation I	Simulation II	Simulation III
Neighbourhood dynamics driver, suitability driver* and zoning driver	✓	✓	✓
Accessibility driver		✓	✓
Differentiation between transport element quality levels			✓
* Note: suitability driver is included, but no rules are set in this particular dataset. See appendix G.			

Table 6.1 - Overview of what is included in the three different simulations

stations. For the roads, the flow is chosen to be the leading traffic characteristic, instead of congestion. An important consideration for using congestion would be that an additional variable, the road capacity, is included. These data are not always reliable. This can be seen in the fact that extreme high I/C ratios are being simulated by Metronamica (for example on ramps). Besides, in the data analysis, the flow level caused more differences in overrepresentation than congestion did. However, congestion levels showed different results than flow levels. Especially the overrepresentation along highly congested roads was far less than high flow roads. To compensate for this effect, while keeping the considerations above in mind, all roads with congestion above 1 are put in a new category, where the attraction is set to zero. Ultimately, the roads of a certain level are divided in three groups (low, medium, high flow), just as in the data analysis. Highly congested roads are a separate category.

6.1.2 Adjustments to Metronamica's local accessibility algorithm

The third simulation requires that roads are classified according to the simulated flow. An alteration had to be made to Metronamica LUT's local accessibility measure in order to incorporate this classification.

The adjusted local accessibility measure differs from the original one in that it chooses other parameter values depending on the flow of the road. The roads are categorised into three flow levels based on their simulated flow (this is done each time step). The user can set different parameters for each flow category. This idea is also implemented in other RIKS models, such as Xplorah (Van Delden, Gutiérrez, Van Vliet, & Hurkens, 2008; Van Delden, Van Vliet, Navarro, & Gutiérrez, 2010).

Note that two roads of the same type (e.g. two highways) which fall into different flow levels are still considered as the same type. Only the nearest of the two is taken into account for the local accessibility calculation. The reason for not including them both, involves practical issues with the distribution of traffic along the road. Consider, for example, a t-junction, where two ends represent the same motorway and the third end represents a lower level road. Because of the third road, the two motorway segments fall into two different flow categories. When modelling different flow categories as different link types, this would lead to double inclusion of the same road. For similar reasons, the road network is usually classified into only a few link types (in the Netherlands application of Metronamica, only two link types are used, despite that the original network data distinguish about eight different levels).

Although the idea behind the local accessibility measure is that when the need to accessibility 'cannot be fulfilled by one link type, then the remaining part can be fulfilled by another link type, and so on' (RIKS, 2011, p. 204), it remains a problem in practice that the more link types are defined, the more uncertainty is introduced with respect to what information is actually added to the equation (do two segments represent different or overlapping roads, etc.). Adding traffic characteristics as separate link types would therefore generate an opposite effect and lead to inaccurate results.

Please refer to appendix I for a technical description of the changes.

Transport element	Quality levels	Transport element	Quality levels
Train stations	Intercity stations Local stations	Level 3 roads	Low flow Medium flow High flow
Level 1 roads	Low flow Medium flow High flow	Level 4 roads	Low flow Medium flow High flow
Level 2 roads	Low flow Medium flow High flow	Highly congested roads	-

Table 6.2 – Transport element quality levels

6.2 Calibration and validation procedure

The period 1989-2000 was used for testing different parameter settings and to obtain a set of final parameter values for the local accessibility of simulation III (calibration). The periods 2000-2006 and 2006-2030 were used to test these parameter values (validation). Both processes are explained here.

6.2.1 Calibration

The calibration (of simulation III) is an iterative procedure in which the model is run for a past period (1989-2000). A set of indicators (explained in section 6.2.3) is used to set parameters in such a way that the fit between data and simulation is improved (Wickramasuriya, et al., 2009). The initial parameter values from earlier calibration efforts by RIKS are taken as a starting point. Based on the data analysis results, some of the parameters are adjusted. For example, the weight of the attraction of commercial land-uses to local stations was made smaller than the attraction to intercity stations. The simulation results are then analysed using the indicators. This is the basis for a further alteration of the parameter values.

Since the data analysis results indicate that urban land-uses show the largest attraction to transport elements (section 5.3), the calibration focuses on parameters that belong to the residential and commercial/industrial land-use types. For the other land-use types, the local accessibility parameter settings are the same in simulation II and III.

Many runs have been carried out to achieve the results below. However, the calibration has not been fully optimised and it needs to be noted that better simulation results can be reached by doing a more comprehensive calibration procedure. Nonetheless, the influence of the majority of the parameters was tested, so the general implications have become clear.

6.2.2 Validation

In the validation, the three simulations are run one more time using the final parameter settings from the calibration step. However, different data is used than in the calibration; in this case a different time period (2000-2006) for the same study area. It is basically tested whether the parameter settings remain useful for other applications than only the one used for calibrating the model. The same indicators are used to assess the validation results.

Additionally, a future validation will be performed. This is a test for the feasibility of the modelled long-term land-use developments. This entails a simulation with 2006 as the start year and 2030 as the end year. Naturally, there is no data available for 2030, so most of the indicators cannot be calculated. However, a visual observation can at least indicate whether the simulated 2030 map appears feasible (note that the long period makes the land-use patterns more sensitive to unrealistic parameter settings). Besides, on the numeric indicators, the cluster size-frequency distribution, can be used to evaluate the long term trend (from 1989 to 2030).

6.2.3 Indicators

Three types of indicators can be distinguished to compare actual and simulated maps in land-use modelling (RIKS, 2008, 2012; White, et al., 1997): a cell-by-cell comparison of the agreement between land-use maps, a more abstract comparison (i.e. land-use patterns) and a visual comparison. A set of indicators, at least one in each category, will be used to assess the similarity between the simulated and actual map, see Table 6.3. The table also gives an overview of what indicators are used in the calibration and validation steps.

6.2.3.1 Indicators for cell-by-cell comparison

Several methods of automated map comparison based on the Kappa-statistics are available. Kappa and Fuzzy Kappa directly compare two land-use maps on a cell-by-cell basis (i.e. data and simulation). The disadvantage of these indicators is that good scores are given when the majority of cells remain the same over the modelling period even though the changes are modelled badly. Therefore, it was chosen to work

with Kappa Simulation (Van Vliet, Bregt, & Hagen-Zanker, 2011) and Fuzzy Kappa Simulation, which are similar methods, but only compare whether land-use changes have been modelled correctly.

Kappa Simulation is a measure of the agreement between two land-use change maps. The actual change map (actual 1989 – actual 2000) and the simulated change map (actual 1989 – simulated 2000) are compared on a cell-by-cell basis. The value for $\kappa_{Simulation}$ lies in $[-1; 1]$ where 1 equals full agreement and 0 equals the same agreement as can be expected with a random allocation of changes (and smaller than 0 equals a worse simulation than random). In practice, values higher than 0.2 are considered to represent a good simulation.

$$\kappa_{Simulation} = \frac{p_0 - p_{e(Transition)}}{1 - p_{e(Transition)}} \quad 6.2$$

Where

$p_{e(Transition)}$ Expected fraction of agreement

p_0 Observed fraction of agreement

Fuzzy Kappa Simulation is a similar measure, but it also takes the neighbourhood of a cell into account. For instance, when a certain change is modelled right next to the cell where it was supposed to be according to the data, it still contributes to the overall score, but in a smaller extent.

6.2.3.2 Indicators for pattern comparison

Three indicators are used to assess land-use patterns. The first is the clumpiness index, which is a measure for the extent to which cells of a given land-use type are connected. The second is the cluster size-frequency distribution, which aims at assessing the patterns of urban areas. The third is an overrepresentation graph as used in the data analysis.

The *clumpiness index* is a single value that expresses to what extent cells of a land-use type are bordering each other (RIKS, 2008). The value lies in $[-1; 1]$ where -1 means minimal connection between cells and 1 means maximum connection. The clumpiness index is used in practice by calculating it for both the actual land-use map and the simulated map (for a given land-use type). Then, a third value is calculated for a map where all land-use changes are randomly reallocated (keeping all non-changed cells the same). This is called a random constraint match (RCM). The value of the simulation should be as close to the value of the actual map as possible, but should at least surpass the value of the RCM.

The *cluster size-frequency distribution* (CSFD) is a measure for the distribution of urban areas (RIKS, 2008). It is especially useful for testing the long-term behaviour of a model. A plot is made with the log of the cluster size on the x-axis and the log of the frequency on the y-axis. It is known from practice that a straight line appears (RIKS, 2008). Plotting this line for a few simulated years, it can be seen whether there is a trend toward more dispersed or more clustered urban areas. The trend should be maintained over the whole calibration period.

Additionally, since this research is about the attraction of infrastructure on land-use developments, the overrepresentation graphs (data analysis; chapter 5) are also used as an indicator. These plot the overrepresentation of land-use types in the neighbourhood of a certain transport element to distance. The graphs of the data analysis are the benchmark. By comparing the overrepresentation graphs of the data to

Indicators	Calibration 1989-2000	Validation 2000-2006	Validation 2006-2030
Cell-by-cell comparison			
Kappa Simulation	✓	✓	
Fuzzy Kappa Simulation	✓	✓	
Pattern comparison			
Overrepresentation graphs	✓	✓	
Clumpiness	✓	✓	
Cluster size-frequency distribution	✓	✓	✓
Visual comparison	✓	✓	✓

Table 6.3 – Indicators overview

those of the simulation, it can be indicated whether to what extent the overrepresentation is modelled correctly. Since the local accessibility parameters include distance decay parameters, changes in these parameters are most easily seen here.

6.2.3.3 Visual comparison

Visual analysis is often a more powerful tool in analysing complex spatial patterns than other techniques (White, et al., 1997). It is used to detect oddities in the simulated maps which are not picked up by any of the numeric indicators. The visual comparison can always be used as a 'veto' to deny the correctness of a simulated map, even though numeric indicators lead to other conclusions.

6.2.3.4 Tool

The Map Comparison Kit (Visser & De Nijs, 2006) is used to calculate the Kappa variants and the clumpiness index. Several available tools were used for visual inspection. The cluster size-frequency distribution was calculated with the RankSizeTool (included in Metronamica installation) and visualised in Excel.

6.3 Results & discussion

The final values of the parameters are shown in Table 6.4. The values of the indicators show some interesting and surprising results. Table 6.5 summarises the numerical indicators for all simulations. The Kappa variants lie between -1 and 1, where values above 0 indicate a better simulation than a random simulation and 1 indicates full agreement between data and simulation. In practice, however, values higher than 0.2 are considered as good simulations. The clumpiness index values given in the table are the difference between the value for the data and for the simulation. Thus, the value should be zero ideally.

Some excerpts from the change maps, which clearly show differences in patterns between simulation I and III, are shown in Figure 6.1 to Figure 6.3 (found from page 66 onward). As an example, the overrepresentation graphs for level 1 roads (motorways) are shown in Figure 6.4 to Figure 6.7. Finally, the cluster size-frequency distributions are shown in Figure 6.8. The simulation results will be discussed below, making extensive use of these tables and graphs.

6.3.1 Simulation I: overrepresentation modelled with neighbourhood dynamics driver

The first point of interest is the overrepresentation graphs for simulation I. The overrepresentation to level 1 roads (motorways) is shown for the data and the simulation in Figure 6.4 and Figure 6.5 respectively. In the

	Simulation II		Simulation III	
	Dist. Decay a_{sf}	Rel. Weight w_{sf}	a_{sf}	w_{sf}
Ramps	10	0.1	10	0.1
Intercity stations	10	0.5	10	0.4
Local stations	10	0.5	8	0.15
Level 1 (low)	8	0.4	8	0.1
Level 1 (medium)	8	0.4	8	0.4
Level 1 (high)	8	0.4	8	0.6
Level 2 (low)	2.5	0.5	2.5	0.5
Level 2 (medium)	2.5	0.5	2.5	0.05
Level 2 (high)	2.5	0.5	2.5	0.2
Level 3 (low)	5	0.8	5	0.8
Level 3 (medium)	5	0.8	5	0.4
Level 3 (high)	5	0.8	5	0.2
Level 4 (low)	5	0.8	5	0.2
Level 4 (medium)	5	0.8	5	0.2
Level 4 (high)	5	0.8	5	0.1
Highly congested	-	0	-	0
Distance decay for waterways and ferries is set to 0				

	Simulation II		Simulation III	
	Dist. Decay a_{sf}	Rel. Weight w_{sf}	a_{sf}	w_{sf}
Ramps	10	0.2	10	0.2
Intercity stations	13	0.3	15	0.3
Local stations	13	0.3	10	0.5
Level 1 (low)	8	0.8	8	0.9
Level 1 (medium)	8	0.8	8	0.7
Level 1 (high)	8	0.8	8	0
Level 2 (low)	4	0.8	4	0.3
Level 2 (medium)	4	0.8	4	0.3
Level 2 (high)	4	0.8	4	0.3
Level 3 (low)	4	0.8	4	0.8
Level 3 (medium)	4	0.8	4	0.8
Level 3 (high)	4	0.8	4	0.3
Level 4 (low)	4	0.8	4	0.8
Level 4 (medium)	4	0.8	4	0.8
Level 4 (high)	4	0.8	4	0.8
Highly congested	-	0	-	0
Distance decay for waterways and ferries is set to 0				

Table 6.4 – Calibrated local access, parameters for residential (left) and commercial/industrial (right)

simulation, the residential overrepresentation is higher for high flow roads, which is in accordance with the actual map. However, the difference between flow levels should be greater according to the data. Commercial developments, however, should be more overrepresented to low flow roads. The simulation with neighbourhood rules only, models the opposite.

This means that some of the relevant patterns are already achieved through setting the neighbourhood dynamics driver. The most relevant question is where these effects should be modelled: in the neighbourhood parameters or in the accessibility parameters. In general, it would be best to model it where the most direct causal relationship is expected (although this is might not be directly clear from the data analysis results).

Modellers, who are charged with the task of calibrating the local accessibility parameters, should therefore always check which of the overrepresentation effects are already present in the simulation without changing any of the local accessibility parameters. The differences between the overrepresentation of the data and the simulation should be leading in adjustment of the parameters.

6.3.2 Simulation II: patterns vs. locations

The differences between simulation I and simulation II tell us something about whether the inclusion of accessibility is beneficial for the model. According to theory and the data analysis, it should be. Simulation I has a larger value for both Kappa Simulation and Fuzzy Kappa Simulation, meaning that it allocates more cells (about) correctly than simulation II. However, looking at the overrepresentation graphs (example of level 1 roads in Figure 6.6) and the clumpiness, simulation II performs slightly better, especially at short distances. The clumpiness index values also indicate a slightly improved pattern.

Thus, the inclusion of accessibility improves the patterns of land-use simulation at the 'cost' of a certain

Simulation	I No accessibility	II With accessibility No differentiation to quality	III With accessibility With differentiation to quality	RCM
Indicator				
Cell-by-cell				
Kappa Simulation total	0.149	0.137	0.121	
Kappa Simulation residential	0.196	0.185	0.142	
Kappa Simulation commercial	0.152	0.143	0.143	
Fuzzy Kappa Simulation*	0.249	0.230	0.214	
Pattern				
Clumpiness residential**	+0.035	+0.031	+0.024	-0.113
Clumpiness commercial**	+0.049	+0.045	+0.026	-0.208
* Radius: 6 cells. Halving distance: 3 cells				
** The value reported here is the difference between the clumpiness index of the data and simulation				

Table 6.5 – Indicator values for calibration (1989-2000) results of both the original and adjusted accessibility measure

Simulation	I No accessibility	II With accessibility No differentiation to quality	III With accessibility With differentiation to quality	RCM
Indicator				
Cell-by-cell				
Kappa Simulation total	0.078	0.076	0.069	
Kappa Simulation residential	0.114	0.108	0.088	
Kappa Simulation commercial	0.069	0.074	0.073	
Fuzzy Kappa Simulation*	0.148	0.139	0.132	
Pattern				
Clumpiness residential**	+0.034	+0.032	+0.027	-0.113
Clumpiness commercial**	+0.028	+0.026	+0.024	-0.208
* Radius: 6 cells. Halving distance: 3 cells				
** The value reported here is the difference between the clumpiness index of the data and simulation				

Table 6.6 - Indicator values for validation (2000-2006) results of both the original and adjusted accessibility measure

number of correctly modelled cells. Although the preference of one indicator over another is always subjective, it needs to be kept in mind that Metronamica's goal is not to perfectly simulate land-use changes, but rather give a good impression of the patterns. In this regard, the inclusion of accessibility (without quality differentiation) can be seen as an improvement.

6.3.3 Simulation III: differentiation in transport element quality

Different weight and distance decay functions are set for different quality levels of the same transport element in simulation III. It can be seen that the clumpiness of the two urban land-uses is closer to the clumpiness of the data map (and significantly closer than the RCM). This effect can also be seen on the map: Figure 6.1 shows that simulation III produces more dispersed commercial patterns than simulation I. Furthermore, Figure 6.2 and Figure 6.3 show that changes due to setting accessibility parameters are very local of nature (i.e. the visual effect is that cells seem to 'move' a little).

The overrepresentation graph of this simulation (Figure 6.7) also approaches the one from the dataset (Figure 6.4) the most. Looking at residential developments, the difference between the flow levels is larger than in simulation I and II (which approaches the data better). For commercial developments, the low flow roads show greater attraction than before and are more similar to the data graphs, although there is still room for improvement. Furthermore, the trend in cluster size-frequency distribution (Figure 6.8) seems the most consistent for simulation III. As said before, this measure is used to test trends over a period of time and the trend should continue in the same qualitative fashion in time (the trend refers to the slope of the line). It can be seen that the slope of lines of simulation III gradually tilts when moving forward in time. The graphs for the other simulations move back and forth. The Kappa values, however, decreased, indicating that the modelling of the land-use change location again worsened.

Concluding, the results of simulation III show that further simulation improvements can be made by setting different parameter values for different quality levels. Just as we saw before, the patterns are improved while fewer locations are modelled correctly.

6.3.4 Implication of choosing the Netherlands as study area

Even though it is shown that the land-use simulations change when the local accessibility parameters are set, the influence of accessibility is more modest than expected. This might be explained by the fact that the Netherlands contains a dense, well-developed infrastructure network that is of consistent quality throughout the country. Therefore, accessibility does not differ greatly spatially. This decreases the influence of accessibility on the transition potential (formula 3.1).

It is expected that when other study areas are examined, setting different parameter values for different quality levels will have a larger beneficial effect on the land-use simulation. This is especially the case for areas with larger spatial differences in the infrastructure network and (therefore) accessibility.

6.4 Conclusions

The goal of this chapter was to test whether the results of the data analysis can be used to improve the land-use modelling in Metronamica. Three simulations have been carried out: the difference between the results of simulation I and II tell us something about the influence of accessibility on the simulation, while the influence of transport element quality differentiation can be derived from the differences between simulation II and III.

Before the simulation results were compared, a calibration was performed. However, since this is a very time-consuming process, the final result of this calibration is by no means optimal, but it is sufficient to analyse what the influences of the inclusion of accessibility and quality differentiation are.

The simulations were compared on a cell-by-cell basis and by pattern. After including accessibility in the simulation, the ability to simulate land-use changes on the correct places decreased, but the pattern of the

simulation approached the data better. Including differentiation to transport element quality strengthened this effect. It is clear that quality differentiation can influence the simulation. In this specific case, increasing pattern indicators could only be achieved at the costs of worsening the cell-by-cell indicators. The inclusion of quality differentiation at least gives the modelled more freedom in achieving the results that are desired for the specific application.

Several factors have to be kept in mind though. Firstly, as said, the calibration has been performed merely roughly, so better results might be reached when more time would be spent. Secondly, the Netherlands might not be the most suited study area, since accessibility is relatively high and constant throughout the area. This makes it harder to reach desired effects (which are usually local of nature) by adjusting the parameters.

6.5 Reflection

6.5.1 Number of roads taken into account in local accessibility

At a certain point in this research, the idea was to include more than one road into consideration in the local accessibility measure. For example, a cell that is close to two different highways is likely to be even more attractive for urban land-use types than a cell that is close to a single highway. It was therefore argued that such a measure would lead to a more realistic estimation of the local accessibility.

However, such an approach was quickly dismissed, mainly due to practical issues in the way the data are provided. On one hand, it is common practice that network maps are built up out of different points or lines, which is especially a problem in the case of a road network. When a cell is close to two road segments, it is hard to determine whether these segments are part of the same road (either overlapping or adjacent) or two different roads. In the latter case, it would also make a difference whether these two road lead to different destinations or not.

On the other hand, while it would be possible to deal with some of these problems to a little extent, it also needs to be taken into account that road maps from different data sources or for different areas are very likely constructed in a completely different manner (in number of overlapping elements, element length, modelling of separate lanes and direction, etc.), making it quite difficult to come up with a generic algorithm that is able to eliminate some of the problems described above.

Concluding, these practical issues with the pre-calculation of network maps would cause the introduction of massive uncertainty, preventing the proposal of a more specific accessibility measure and confirming that including one road per type is definitely the most ideal solution when trading off between introducing more errors and theoretical aspects.

6.5.2 Random coefficient

The simulations in this chapter were done without a random coefficient in the transition potential formula (formula 3.1). Some tests were done including this factor, which is known from experience to generate better simulation. However, the calibration should be redone to optimise for the inclusion of this effect. This is not done here due to time constraints. The probability maps of the monte carlo analysis (random factor: 0.2; iterations: 6) are included in appendix J.

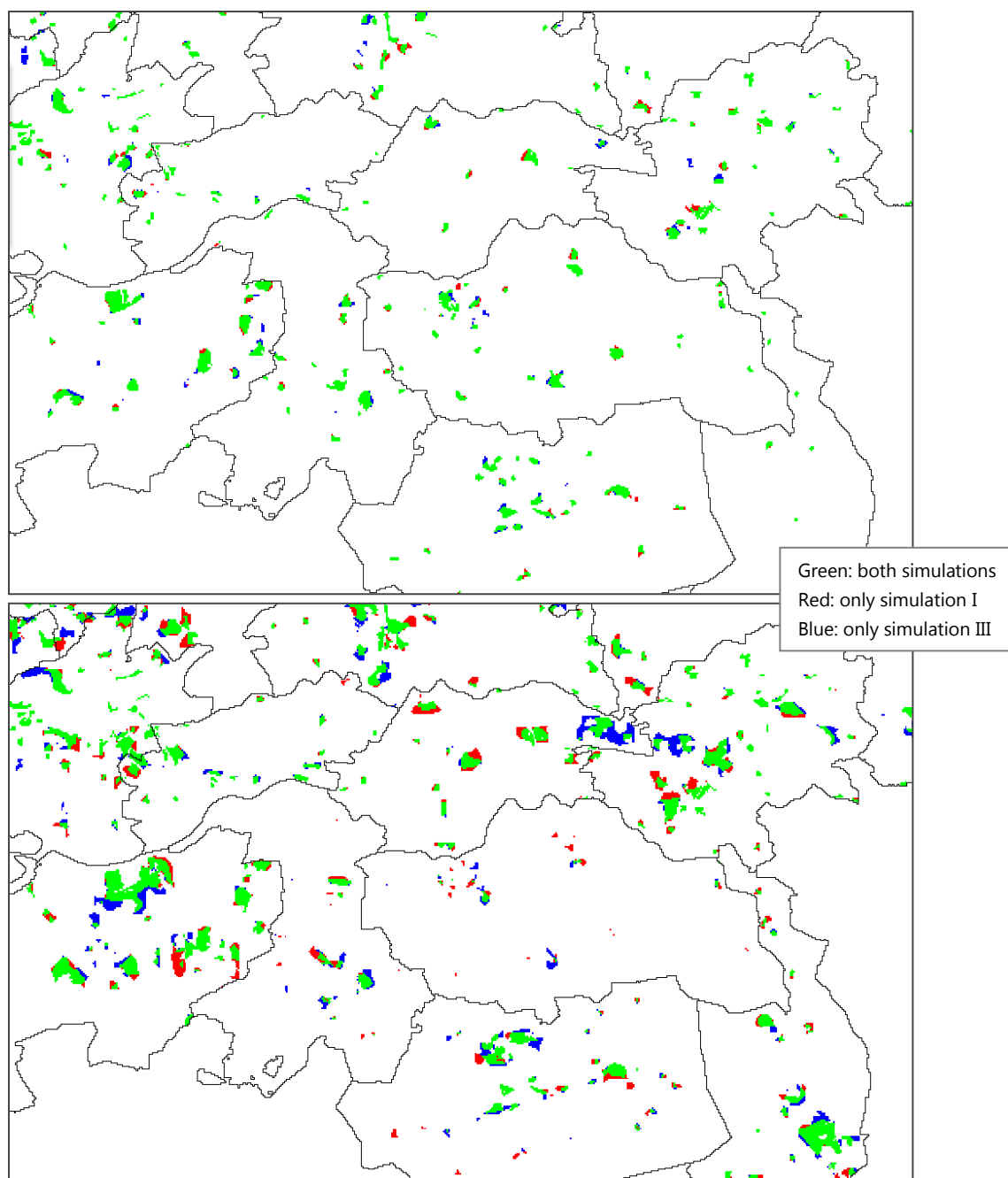


Figure 6.1 - Commercial land-use cells in simulation I and III in the province of Noord-Brabant. Top: 2000; bottom: 2030.

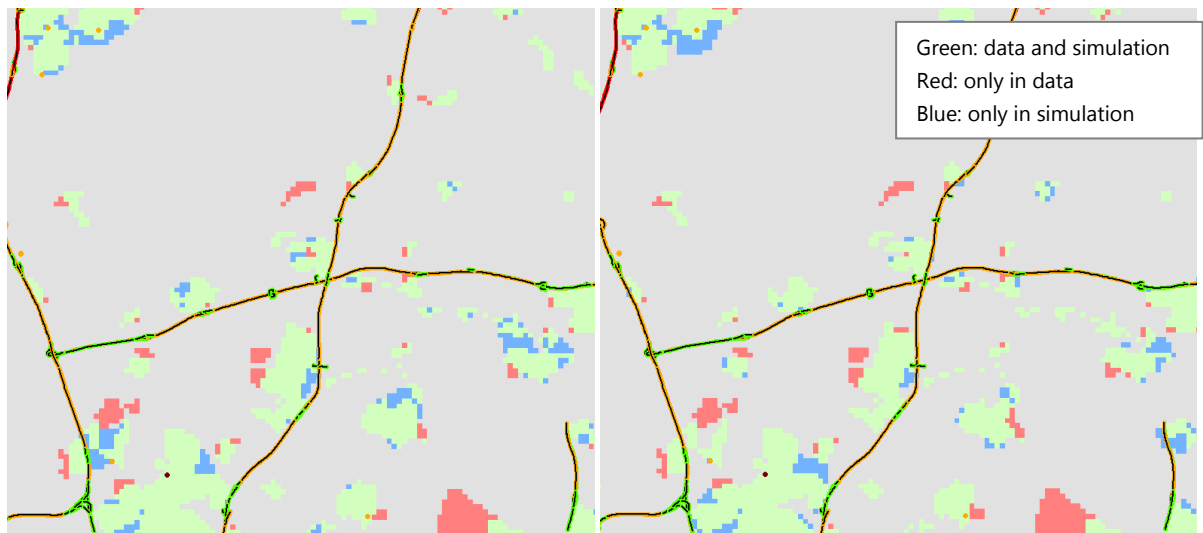


Figure 6.2 – Residential land-use simulation and data (2000). Left: sim. I; right: sim. III (excerpt: western part of N-Brabant)



Figure 6.3 - Commercial land-use simulation (2000) and data. Left: sim. I; right: sim. III (excerpt: western part of N-Brabant)

See appendix K for full maps.

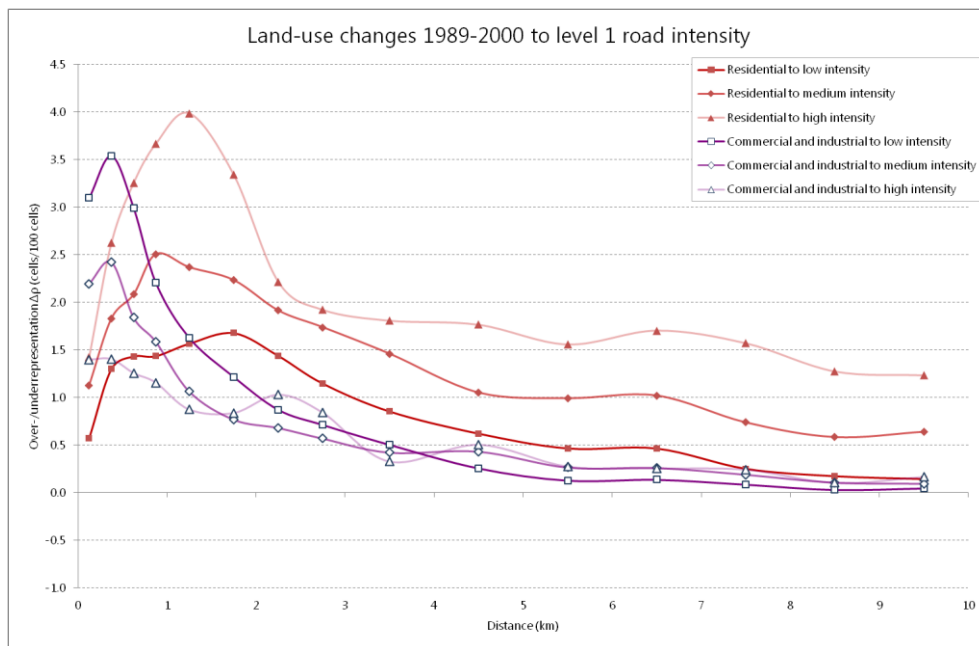


Figure 6.4 - Overrepresentation graphs for level 1 roads (motorways) in the data

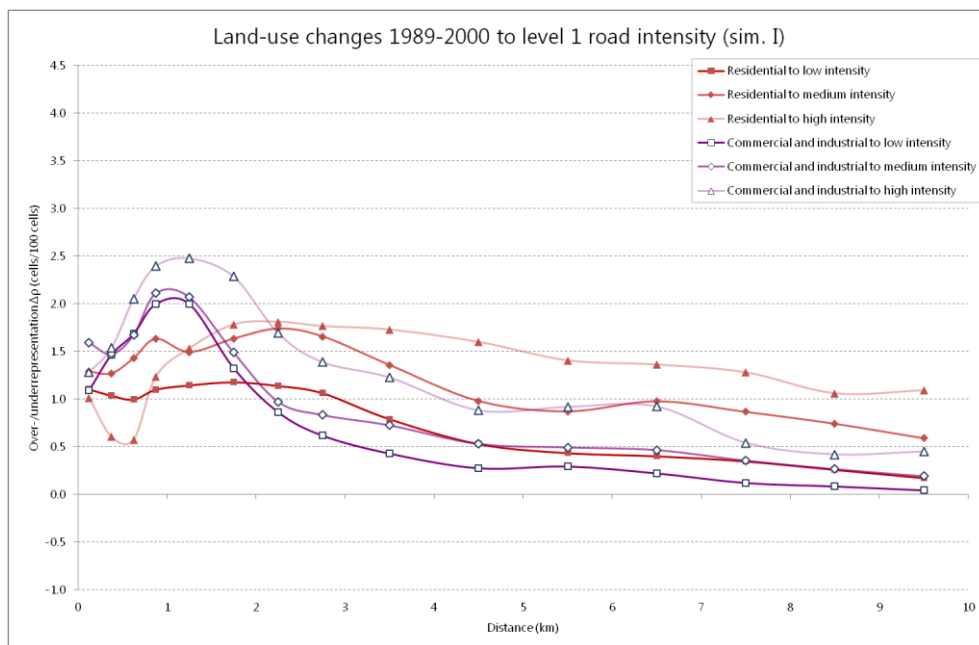


Figure 6.5 – Overrepresentation graphs for level 1 roads (motorways) in simulation I

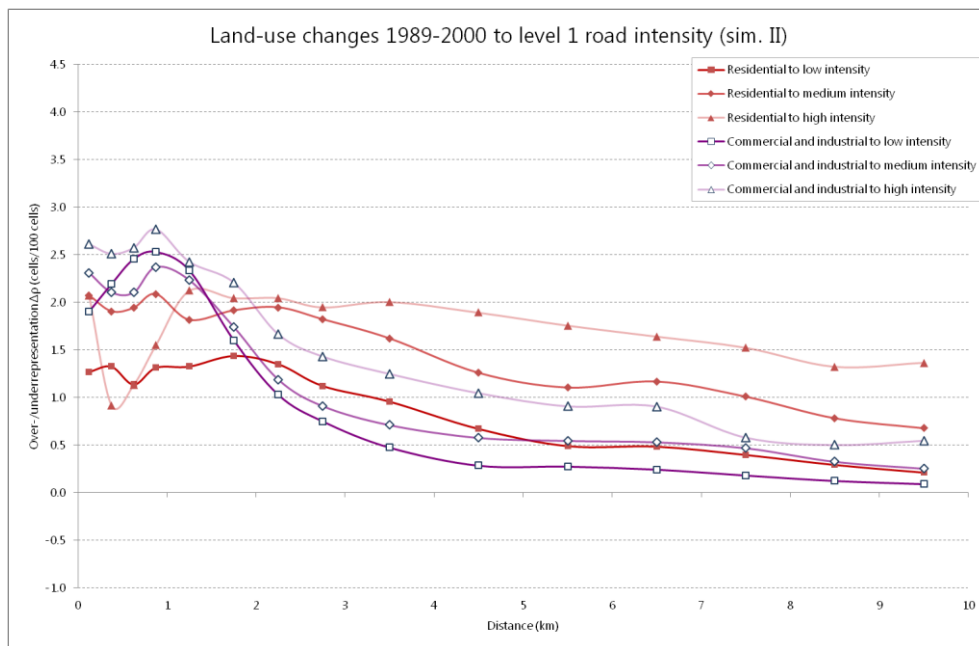


Figure 6.6 - Overrepresentation graphs for level 1 roads (motorways) in simulation II

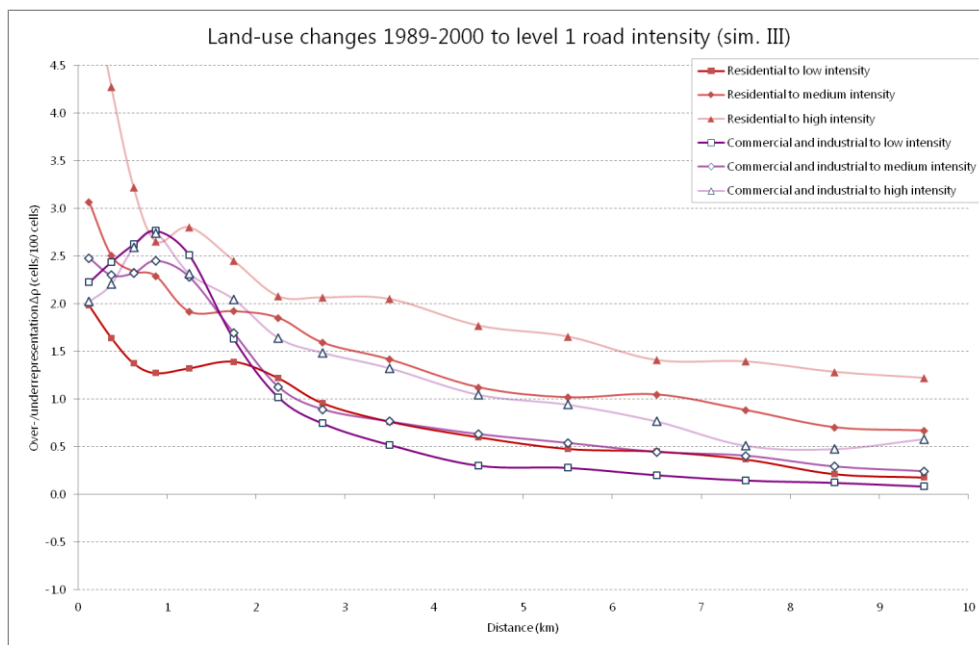


Figure 6.7 - Overrepresentation graphs for level 1 roads (motorways) in simulation III

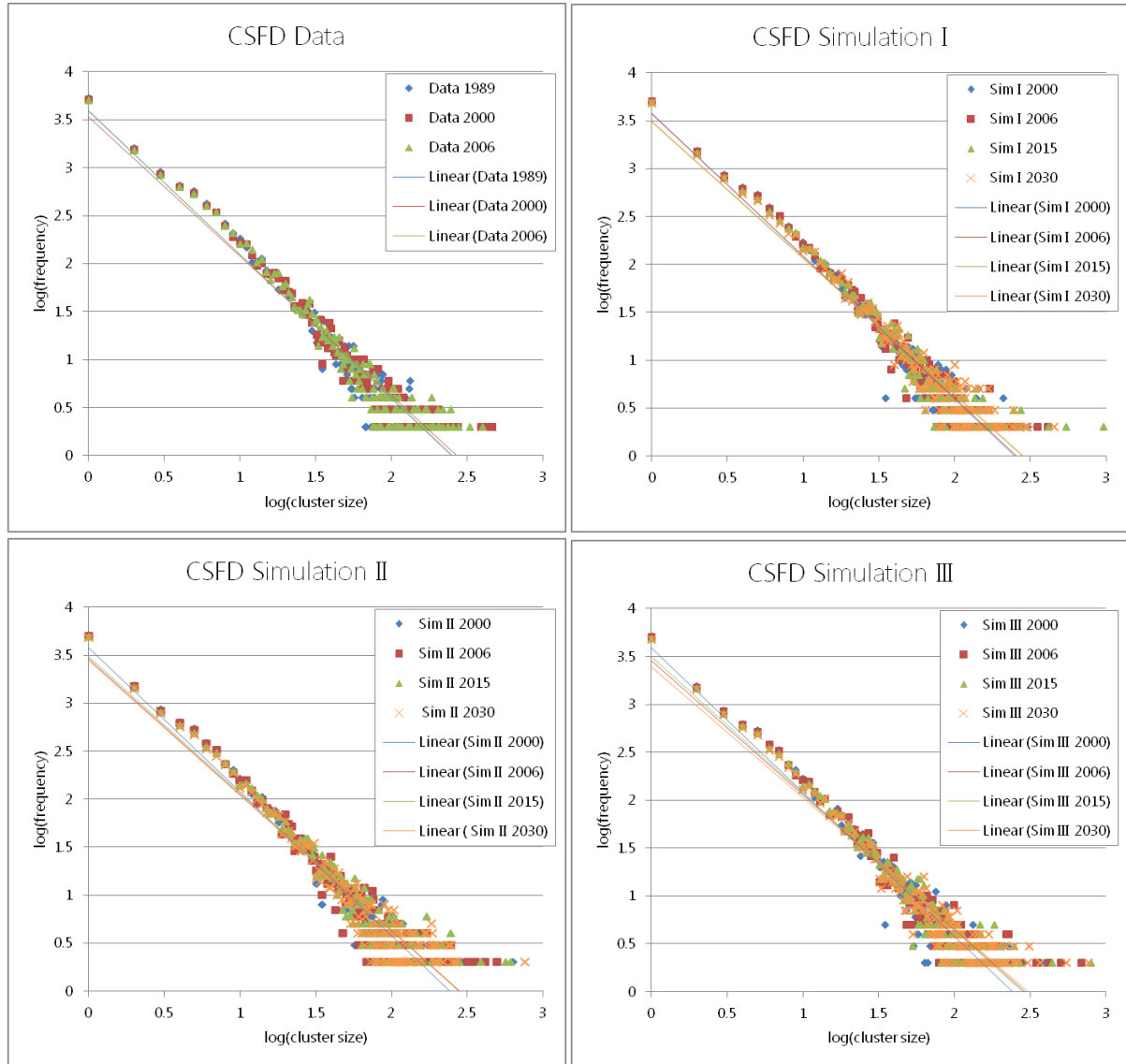


Figure 6.8 – Cluster size-frequency distributions for the data and the three simulations. The simulation results for 2006 were obtained with a simulation with start year 2000 (using 2000 data) and the simulation results for 2015 and 2030 were obtained with a simulation with start year 2006 (using 2006 data).

7 Conclusion

The goal of this research was to find a way to improve the land-use modelling in Metronamica by conducting an empirical research and testing how the results can be used in Metronamica. The focus was on the influence of transport elements on nearby land-use developments, which is modelled in Metronamica in the local accessibility driver.

An empirical research was done focussing on the overrepresentation of land-use developments near transport elements using data for the Netherlands of the period 1989-2000. Land-use developments were defined as cells where the land-use changed during this period, where the new land-use type is considered the 'value' of the land-use development. The results of the empirical research were then used in Metronamica to test the influence of the inclusion of accessibility and, more importantly, the inclusion of differentiation between transport element quality levels on the simulation.

7.1 Research question 1

What is the role of transport elements and accessibility in explaining land-use changes in the Netherlands?

The first research question was answered by means of a data analysis in chapter 5. Prior research of Verburg et al. (2004) already indicated that accessibility is one of the most important factors in explaining land-use development in the Netherlands in recent decades. The data analysis in this research investigated this relation in more depth. Indeed, interesting overrepresentations were found.

As expected in theory (e.g. Wegener & Fürst, 1999), urban land-use types (residential and commercial & industrial areas) are attracted to transport elements in general. However, they also prefer specific quality levels. For example, commercial developments appear more along low flow motorways than high flow motorways. In general, the findings were as expected in theory and as previously indicated in empirical research. One notable exception is the overrepresentation close to train stations, which is clearly present in the overrepresentation graphs, but has far less back-up from existing research than the other relations found.

It is important to stress that not all the overrepresentation results can be explained and, hence, no direct causal relationship is present in all cases. This is, for example, the case for the overrepresentation of artificial nature near transport elements. In reality, it is expected that artificial nature arises where urban areas are developing, independent from whether a transport element is close (but the urban developments are indeed attracted to the transport elements).

7.2 Research question 2

How can accessibility modelling be improved in Metronamica?

Accessibility modelling can be improved by introducing differentiation between transport element quality levels, since the data analysis clearly indicates different overrepresentation patterns for different quality levels. Dynamic (i.e. simulated) road quality characteristics (flow, congestion) should be used for the best results. The flow would be the most suited characteristics, since it does not depend on an extra variable (capacity). However, there are important differences between flow and congestion in the data analysis, so ignoring congestion might not lead to the best results. For this research, it was chosen to work with flow for defining the road's quality levels, but highly congested roads were put into a separate category.

Furthermore, the data analysis results can be used to set local accessibility parameters in Metronamica. The focus should be on the overrepresentation that is expected to have a causal relationship. The modelling of other effects might benefit the overall simulation results as well, although the danger arises that the wrong processes are being modelled. In the end, these judgements are partly subjective and should be made by the modeller with careful consideration.

7.3 Research question 3

Do the suggestions from question 2 improve the simulation performance?

Using the data analysis results to set the local accessibility parameter values, different simulations were carried out and compared together with a set of indicators. The 'improvement' was measured with a set of indicators. Some of these indicators are measures for how well the modelled location of land-use changes fit the data. Others are more abstract measures that evaluate how well the land-use patterns fits the data.

It was found that including local accessibility in the simulation mainly improved the modelling of land-use patterns, while the number of correctly modelled land-use developments decreased. Including different parameter values for different quality levels of the same transport element, strengthens this effect.

Whether this can be described as 'an improvement to Metronamica's modelling', as stated in the research goal, remains subjective. Metronamica's nature, however, is not to perfectly simulate land-use changes, rather to understand the underlying processes that express themselves in land-use patterns. In any case, differentiation between transport element quality levels at least gives the modeller more opportunities to find a balance between patterns and location. Where this balance exactly is, depends on both the application and the modeller.

7.4 Discussion

The use of this research is twofold. The interesting overrepresentation patterns found in the data analysis might be of interest to any land-use modeller or researcher. The results can also be used for other applications than Metronamica. It is also shown that the land-use simulation can be further refined after introducing transport element quality levels. This improvement might, however, be larger for other study areas with more spatial variation in infrastructure quality.

Although the results of the research are promising, some remarks need to be made. These were discussed in the relevant chapters, but the most important remarks are repeated here. Firstly, the calibration was only done roughly due to its complexity and time constraints. A more comprehensive calibration might lead to better parameter settings and to further insights of the influence of transport element quality on the simulation. This research provides first insights, but it would be too much to state that this research contains all there is to understand. Secondly, assumptions about causality in the overrepresentation graphs should be re-evaluated when further insights in the complex land-use developments arise. Although the overrepresentation graphs lead to interesting patterns, care must be taken in interpretation and use.

Interesting topics to expand on this research could either focus on repeating the data analysis for different study areas or on further investigating the effects of the local accessibility parameters on the simulation.

The advantage of using the Netherlands as a study area is that many detailed data are available. However, in retrospect, the downside is that the Netherlands has a very well-developed, dense and consistent transport network over the country. The differences from one region to the other are relatively small compared to other countries. Therefore, attraction effects of transport elements to land-use developments might be more obvious (or perhaps different) in other study areas. It is advised to choose a study area that has different characteristics than the Netherlands, especially concerning the state of the transport network. Culture, social economic factors will influence the results as well. Comparing overrepresentation graphs for another area to the ones for the Netherlands, will surely lead to interesting insights.

As stated, the calibration of the Netherlands was done roughly. It is very likely that a further calibration will improve the simulation. Even the cell-by-cell indicators might become better when an optimal set of parameter settings is used. For Metronamica modellers and users, it would be of great value to further investigate the implications of settings different weights for different quality levels on the simulation, and specifically whether it is possible to improve both cell-by-cell and patterns indicators at the same time.

8 References

- Borzacchiello, M. T., Nijkamp, P., & Koomen, E. (2010). Accessibility and urban development: a grid-based comparative statistical analysis of Dutch cities. *Environment and Planning B: Planning and Design*, 37(1), 148-169.
- Bruinsma, F. R. (1994). *De invloed van transportinfrastructuur op ruimtelijke patronen van economische activiteiten*. Vrije Universiteit, Amsterdam.
- Bürgi, M., & Turner, M. G. (2002). Factors and processes shaping land cover and land cover changes along the Wisconsin River. *Ecosystems*, 5(2), 184-201.
- De Bok, M. (2007). *Infrastructure and firm dynamics: a micro-simulation approach*. Delft University of Technology, Delft.
- De Bok, M., & Sanders, F. (2005). Firm location and the accessibility of locations: empirical results from The Netherlands. *Transportation Research Record*, 1902, 35-43.
- De Nijs, T. (2009). *Modelling land use change: improving the predictions of future land use patterns*. Universiteit Utrecht, Utrecht.
- Densham, P. J. (1991). Spatial decision support systems. *Geographical information systems. Vol. 1: principles*, 403-412.
- European Environment Agency. (n.d.). Corine land cover maps. Retrieved 2012 <http://www.eea.europa.eu/data-and-maps>
- Geurs, K. T., & Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography*, 12, 127-140.
- Geurs, K. T., & Van Wee, B. (2006). *Ex Post* Evaluation of Thirty Years of Compact Urban Development in the Netherlands. *Urban Studies*, 41(1), 139-160.
- Hansen, W. G. (1959). How accessibility shapes land use. *Journal of American Institute of Planners*, 25(1), pp. 73-76.
- Iacono, M. J., & Levinson, D. M. (2008). Predicting land-use change: how much does transport matter? *Transportation Research Record* 2119, 130-136.
- Leitham, S., McQuaid, R. W., & Nelson, J. D. (2000). The influence of transport on industrial location choice: a stated preference experiment. *Transportation Research Part A: Policy and Practice*, 34(7), 515-535.
- Levinson, D. (2008). Density and dispersion: the co-development of land use and rail in London. *Journal of Economic Geography*, 8, 55-77.
- Levinson, D. (2011). The Coevolution of Transport and Land Use: An Introduction to the Special Issue and an Outline of a Research Agenda. *Journal of Transport and Land Use*, 4(2), 1-3.
- Levinson, D., & Xie, F. (2011). Does First Last? The Existence and Extent of First Mover Advantages on Spatial Networks. *Journal of Transport and Land Use*, 4(2), 47-69.
- Morris, J. M., Dumble, P. L., & Wigan, M. R. (1978). Accessibility indicators for transportation planning. *Transportation Research A*, Vol. 13, pp. 91-109.
- Naus, N. J. H. (2009). *Understanding neighbourhood effects of land use change to improve the calibration procedure of a CA-based land use model*. Wageningen University and Research Centre, Wageningen.
- Nirov. (2010). De Nieuwe Kaart van Nederland. Retrieved May 2012 http://www.nirov.nl/Home/Projecten/De_Nieuwe_Kaart/Website/Producten.aspx
- Ortúzar, J. d. D., & Willumsen, L. G. (2001). *Modelling transport; 3rd edition*. Chichester, UK: Wiley.
- Polzin, S. E. (1999). Transportation/Land-Use Relationship: Public Transit's Impact on Land Use. *Journal of Urban Planning and Development*, 125(4), 135-151.
- Rijkswaterstaat. (n.d.). Dataportal. Retrieved 2012 www.dataportal.nl
- RIKS. (2008). Assessment and scenarios of land use change in Europe. Maastricht.
- RIKS. (2011). *Metronamica documentation*.
- RIKS. (2012). [Presentation/workshop] RIKS calibration course. Maastricht, the Netherlands.
- Ripley, B. D. (1979). Tests of 'Randomness' for Spatial Point Patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(3), 368-374.
- Van Delden, H. (2010). [Lecture presentation] Decision support systems and integrated spatial modelling.
- Van Delden, H., Gutiérrez, E., Van Vliet, J., & Hurkens, J. (2008). Xplorah, A multi-scale integrated land use model. In M. Sánchez-Marré, J. Béjar, J. Comas, A. Rizzoli & G. Guariso (Eds.), *Proceedings of the iEMSs Fourth Biennial Meeting: Integrating sciences and information technology for environmental assessment and decision making*. International Environmental Modelling and Software Society (pp. 827-834). Barcelona, Spain.
- Van Delden, H., & Hurkens, J. (2011). *A generic Integrated Spatial Decision Support System for urban and regional planning*. Paper presented at the MODSIM11 International Congress on Modelling and Simulation.

- Van Delden, H., Van Vliet, J., Navarro, C., & Gutiérrez, E. (2010). *The Xplorah SDSS - Supporting integrated planning on the island of Puerto Rico*. Paper presented at the 10th International Conference on Design & Decision Support Systems in Architecture and Urban Planning, Eindhoven, The Netherlands.
- Van Vliet, J., Bregt, A. K., & Hagen-Zanker, A. (2011). Revisiting Kappa to account for change in the accuracy assessment of land-use change models. *Ecological Modelling*, 222(8), 1367-1375.
- Verburg, P. H., de Nijs, T. C. M., Ritsema van Eck, J., Visser, H., & de Jong, K. (2004). A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, 28(6), 667-690.
- Verburg, P. H., Ritsema van Eck, J. R., de Nijs, T. C. M., Dijst, M. J., & Schot, P. (2004). Determinants of land-use change patterns in the Netherlands. *Environment and Planning B: Planning and Design*, 31(1), 125-150.
- Visser, H., & De Nijs, A. C. M. (2006). The Map Comparison Kit. *Environmental Modeling and Software*, 21, 346-358.
- Wegener, M. (2004). Overview of land-use transport models. In J. B. D. Hensher, K. Haynes and P. Stopher (Ed.), *Handbook of Transport Geography and Spatial Systems*: Elsevier
- Wegener, M., & Fürst, F. (1999). Land-use transport interaction: state of the art *TRANSLAND Deliverable D2a*. Dortmund: University of Dortmund.
- White, R., Engelen, G., & Uljee, J. (1997). The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design*, Volume 24, 323-343.
- Wickramasuriya, R. C., Bregt, A. K., Van Delden, H., & Hagen-Zanker, A. (2009). The dynamics of shifting cultivation captured in an extended Constrained Cellular Automata land use model. *Ecological Modelling*, 220(18), 2302-2309. doi: 10.1016/j.ecolmodel.2009.05.021
- Wiegand, T., & Moloney, K. (2004). Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, 104(2), 209-229.
- Zondag, B. (2007). *Joint modeling of land-use, transport and economy*. Delft University of Technology, Delft.

Appendix A: Data description

Land-use maps

Type: raster (.rst with auxiliary .rdc)

Source: Corine

The land-use maps were originally retrieved from the Corine project (European Environment Agency, n.d.), which provides free to use land-use maps for European countries. These maps were edited for use in Metronamica before the start of this research. The main change is that some land-use classes were merged. The exact transformation will not be discussed here, as the Metronamica maps are seen as the input maps of this research.

The land-use maps are raster maps with a resolution of 250x250 m. The range of pixel values and their meanings are given in Table 0.1.

Metronamica road map

Type: shapefile (.shp with auxiliary .dbf, .sbn, .sbx, .shx)

Course: NWB

Used for: road type, simulated flows, simulated congestion, simulated speed

The road file is a shapefile (.shp) that exclusively contains polylines. Each line represents a road segment. Table 0.2 shows the incorporated attributes for every link.

The 'roadtype' attributes distinguishes each road by type, where the coding of Table 0.3 is used. The 'intens', 'congest', and 'speed' attributes are reserved for simulated flow, congestion and speed values. The numbers 01, 02 and 03 respectively stand for morning peak, evening peak, and whole day.

Pixel value	Land-use categories
0	Other agricultural areas
1	Pastures
2	Residential areas
3	Commercial and industrial areas
4	Artificial green
5	Forests
6	Nature
7	Wetlands
8	Fresh water
9	Marine water
10	Infrastructure
11	Mines and dumpsites
12	Port areas
13	Airports

Table 0.1 – Pixel values in the land-use raster maps

Road type attributes				
Link_id	Weefvakken	VeldA	Gem_naam	Intens_03
Segment_ID	Grens	VeldB	Prv_code	Speed_01
Linktype	Signal	VeldC	Prv_naam	Speed_02
Wegsoort	NWBstatus	NOMO1	Commertaar	Speed_03
Afstand	Stt Naam	NOMO2	LinkNr	Congest_01
Geb_code	WegBehSrt	H_int	Roadtype	Congest_02
H_sel	H_lanes	Verhard	AccType	Congest_03
H_snel_wet	HWN	H_Baan	ExtraCost	IntensMax
H_snel_mod	Jaar	Doelstrook	Preintens	CongestMax
H_cap	MIT_cat	MengType	Modified	SpeedMin
H_stroken	KnoopA	Wegnummer	Intens_01	
Ex_stroken	KnoopB	Gem_code	Intens_02	

Table 0.2 – Attributes in the road type shapefile

Field value	Meaning
0	Virtual roads (created in Metronamica for calculation purposes)
1	Highways
2	Main roads
3	Access roads (closed declaration)
4	Access roads (mixed traffic)
5	Private roads
6	City access roads
7	Neighbourhood access roads
8	Ferries

Table 0.3 –Field values of the ‘roadtype’ attribute

Rijkswaterstaat road map

Type: shapefile (.shp with auxiliary .dbf, .sbn, .sbx, .shx)

Used for: observed flows

The Dutch Ministry of Infrastructure and the Environment provides an online database in which data on the Dutch road network are stored (Rijkswaterstaat, n.d.), which was used to retrieve data on observed flows for 1989, 1999 and 2006. (Note: these data were actually retrieved via e-mail conversations with Rijkswaterstaat staff, since the portal itself was offline due to maintenance during this research.)

Important to note is that this road network only includes a portion of the roads in the road type map. The overlap between the two sets is perfect (i.e. the networks have the same accuracy).

Table 0.4 contains a list of all included attributes. The attributes ‘werkdag’ and ‘weekdag’ contain the average daily flows for this link for weekdays and all days respectively.

Observed road flow attributes			
Wvk_id	Wegnummer	Gme_id	Beginkm
Wvk_begin	Wegdeelltr	Gme_naam	Eindkm
Baannummer	Stt_naam	Beginafstand	Werkdag
Wegbersrt	Wpsnaamnem	Eindafstand	Weekdag

Table 0.4 – Attributes in the observed road flow shapefile

Appendix B: Road classification intervals

		Level 1	Level 2	Level 3	Level 4
Flow	Low	[0; 1500]	[0; 500]	[0; 600]	[0; 500]
	Medium	(1500; 3600]	(500; 1200]	(600; 1400]	(500; 1300]
	High	(3600; 9932]	(1200; 3257]	(1400; 4678]	(1300; 3669]
Congestion	Low	[0; 0.31]	[0; 0.22]	[0; 0.29]	[0; 0.33]
	Medium	(0.31; 0.67]	(0.22; 0.49]	(0.29; 0.61]	(0.33; 0.72]
	High	(0.67; 2.39]	(0.49; 1.21]	(0.61; 2.60]	(0.72; 2.92]
Operating speed	Low	[0; 70]	[0; 60]	[0; 40]	[0; 20]
	Medium	(70; 105]	(60; 90]	(40; 65]	(20; 40]
	High	(105; 120]	(90; 120]	(65; 80]	(40; 50]

Note: the interval boundaries were determined by Jenks natural break classification method. This method is based on classifying groups in such a way, that the variance within the groups is minimised while the variance between the groups is maximised. A tool available in ArcMap was used for this calculation.

Appendix C: Region maps



Regions map (COROP)



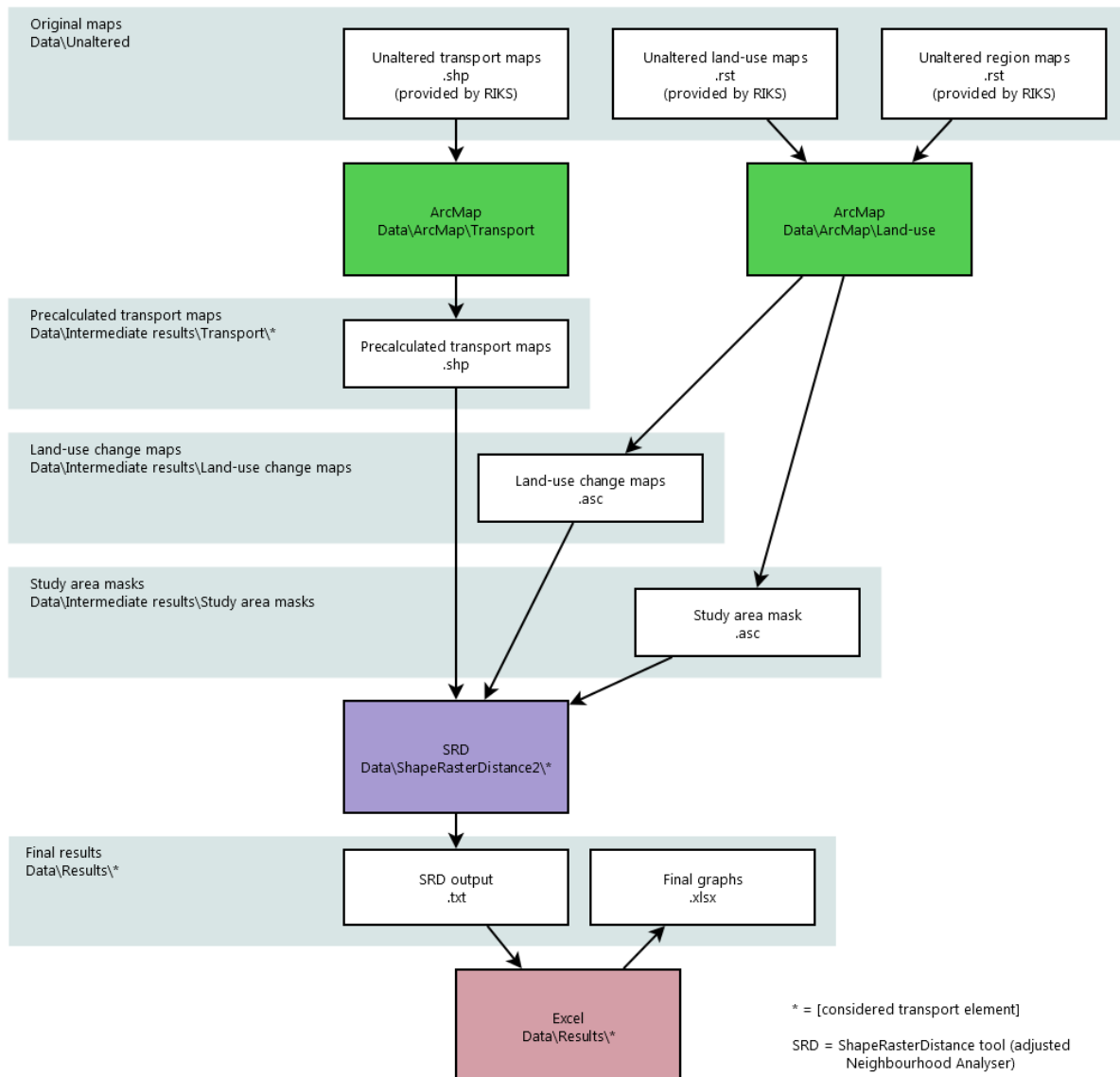
Transport zones map

This is an example of the Excel-sheet that was used to create the overrepresentation graphs. The settings tab is used to manually input some characteristics of the input data. These influence both the calculation and automatically create the correct labels in the graphs. The calculation sheet is basically the Δp calculation as found in section 5.1.3.

Transport element type	Interval categories	
Point	Lower	0 1 2 3 4 6 8 10 12 16 20 24 28 32 36
	Upper	1 2 3 4 6 8 10 12 16 20 24 28 32 36 40
Transport element name	Transport element length (m)	
Stations by type	Only relevant for line transport elements	
Land-use element name	9892185	
land-use changes 1989-2000	2019301	
	12479089	
Study area size (cells)	6090283	
564.581		
Cell size (m)		
250		
Title y-axis		
Over-/underrepresentation Δp (cells/100 cells)		
Title x-axis		
Distance (cells)		

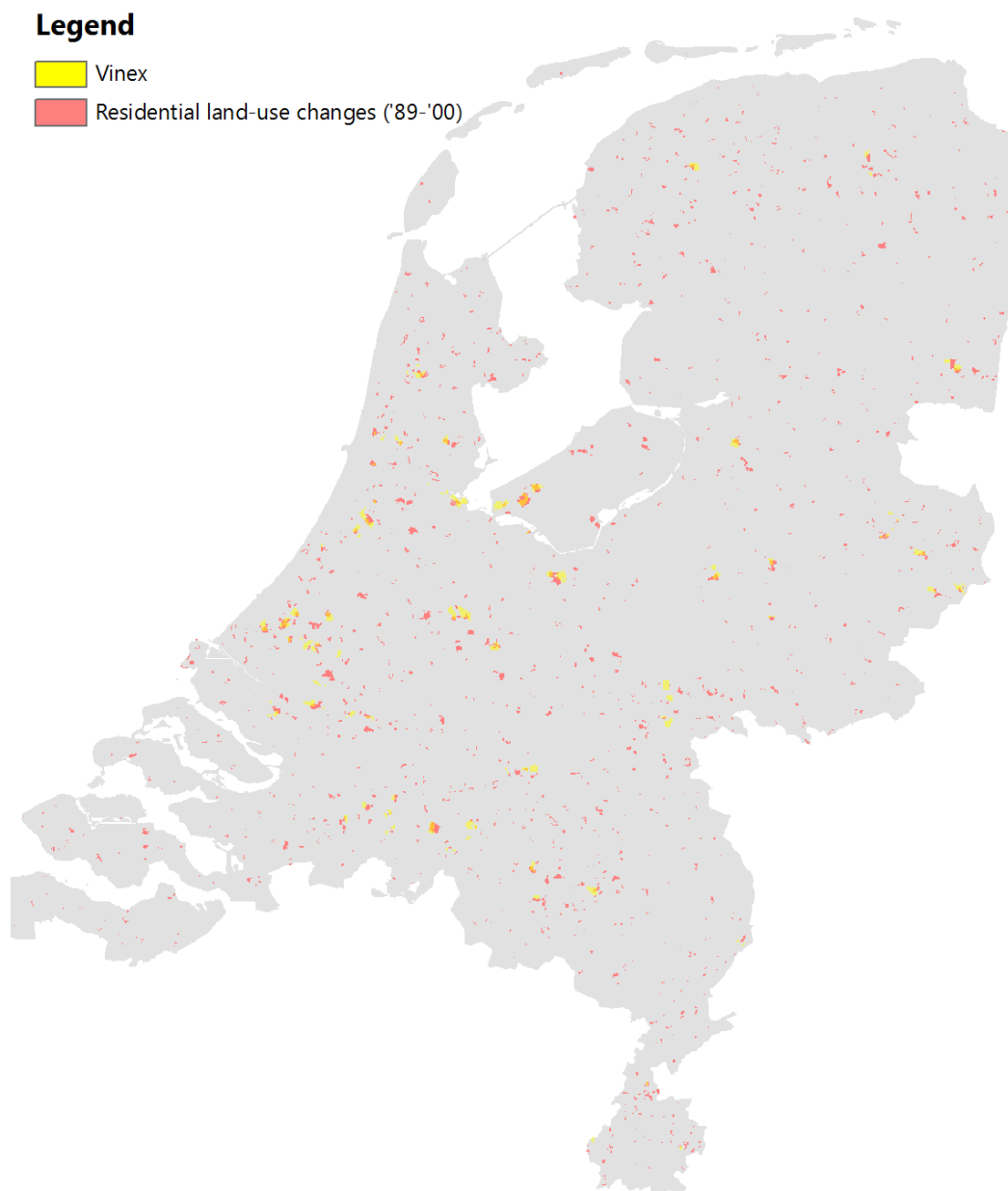
AccType	5	Adjusting this filter value will automatically change the filter of all other Pivot tables in ti																			
Sum of Count	Column Labels																				
Row Labels	[0, 1]	[1, 2]	[2, 3]	[3, 4]	[4, 6]	[6, 8]	[8, 10]	[10, 12]	[12, 16]	[16, 20]	[20, 24]	[24, 28]	[28, 32]	[32, 36]	[36, 40]	Grand Total					
0		0	0	0	0	0	6	23	108	141	180	188	318	366	429	1759					
1				0	0	0	2	1	2	0	13	17	13	30	25	43	146				
2		0	3	17	22	73	143	253	366	754	821	714	708	640	671	860	6045				
3		4	7	12	12	70	153	173	202	469	336	317	267	253	277	274	2826				
4		0	1	2	9	30	25	59	93	179	159	165	285	256	162	231	1656				
5		0	0	0	0	0	3	6	18	43	83	97	136	159	125	109	779				
6						0	0	0	3	29	28	45	47	46	45	74	317				
Grand Total		4	11	31	43	173	326	498	707	1582	1581	1535	1644	1702	1671	2020	13528				
Lower limit		0	1	2	3	4	6	8	10	12	16	20	24	28	32	36					
Upper limit		1	2	3	4	6	8	10	12	16	20	24	28	32	36	40					
Observ. Area		3	9	16	22	63	88	113	138	352	452	553	653	754	855	955					
Land-use	Count	Auto-legend values						Trans.	Count	I (m)	I (cells)	I (cls/el)	Tr. Elem. Type								
0		4250	Other agriculture					5	60	9892185	39568,7	659,479	Point								
1		329	Pastures					6	367	2019301	8077,2	22,0087									
2		7854	Residential					999	0	1,2E+07	49916,4	#DIV/0!	Tr. Elem. Name								
3		3809	Commercial and indust					999	0	6090283	24361,1	#DIV/0!	Stations by type								
4		1997	Artificial green					999	0	0	0	#DIV/0!									
5		1790	Forests					999	0	0	0	#DIV/0!	L.u. Elem. Name								
6		1494	Nature					999	0	0	0	#DIV/0!	land-use changes 1989-2000								
								999	0				Study area size								
								999	0				564581 cells								
								999	0				Cell-size								
								Current	60		659,479		250 m								
		0,5	1,5	2,5	3,5	5	7	9	11	14	18	22	26	30	34	38					
Other agriculture		-0,75	-0,75	-0,75	-0,75	-0,75	-0,75	-0,66	-0,48	-0,24	-0,23	-0,21	-0,27	-0,05	-0,04	0,00					
Pastures		-0,06	-0,06	-0,06	-0,06	-0,06	-0,02	-0,04	-0,03	-0,06	-0,01	-0,01	-0,03	0,01	-0,01	0,02					
Residential		-1,39	-0,86	0,41	0,28	0,55	1,32	2,34	3,02	2,18	1,63	0,76	0,41	0,02	-0,08	0,11					
Commercial and in		1,45	0,56	0,60	0,23	1,18	2,22	1,87	1,76	1,55	0,56	0,28	0,01	-0,12	-0,13	-0,20					
Artificial green		-0,35	-0,18	-0,14	0,33	0,44	0,12	0,52	0,77	0,49	0,23	0,14	0,37	0,21	-0,04	0,05					
Forests		-0,32	-0,32	-0,32	-0,32	-0,32	-0,26	-0,23	-0,10	-0,11	-0,01	-0,02	0,03	0,03	-0,07	-0,13					
Nature		-0,26	-0,26	-0,26	-0,26	-0,26	-0,26	-0,26	-0,23	-0,13	-0,16	-0,13	-0,14	-0,16	-0,18	-0,14					
		0,5	1,5	2,5	3,5	5	7	9	11	14	18	22	26	30	34	38					
Other agriculture																					
Pastures																					
Residential																					

Appendix E: Flowchart of data analysis process



- **SRD** refers to the ShapeRasterDistance tool, which is provided by RIKS. It is similar to the Neighbourhood Analyser and counts the number of land-use elements in proximity of transport elements (given a certain distance range).
 - Input:
 - Land-use (change) map (.asc or .rst)
 - Transport network map (.shp)
 - Study area mask (.asc or .rst)
 - Min. and max. distance from transport element (definition observation area) (#cells)
 - Output:
 - List of cells counted within observation area (.txt)
- The **unaltered land-use maps** were edited by RIKS and are not the original Corine maps
- **Pre-calculated transport maps**: contain transformations for making transport map ready for use in SRD. For example: in the ramp map, ramps at highways were indicated by two points (one at each side of the road). This was corrected to one point.
- **Study area mask**: a file that is used by SRD to identify which cells of the (square) raster maps belong to the study area

Appendix F: Vinex map



Appendix G: Implication of overlapping observation areas

This appendix mathematically describes the error in the data analysis due to using estimated values for the observed land-use elements instead of the actual values. The overrepresentation method was defined as follows earlier:

$$\Delta\rho_c = 100 * \frac{\sum_{x=1}^{n_{tr}} n_{lu,observed,x}}{n_{tr}A_{oa,c}} - 100 * \frac{n_{lu}}{A_{sa}} \quad 0.1$$

The idea behind the left term this function is that the density of all land-use elements within all observation areas is retrieved. Although it might seem this exact density is calculated, two errors are present in this calculation:

- The actual observation area is smaller than the estimation in the denominator. This is caused by the fact that the observation areas can overlap and that the overlapping areas are included multiple times.
- The actual number of land-use elements in all observation areas is smaller than the estimation in the numerator. This is caused by the fact that land-use elements in overlapping observation areas are counted multiple times.

The reason why the actual observation area and actual number of points cannot be used in the formula is that the calculation algorithms in the available tools do not allow for this option without massive alteration (this is especially the case for the second issue in combination with the SRD tool).

The implication of this error will be discussed below. The conclusion will be that these two errors completely negate each other **only** if assumed that multiple nearby transport elements do **not** increase/decrease the attraction compared to the situation where only one nearby transport element is present. These relations are, however, expected, and therefore, when analysing any of the data analysis results, it has to be kept in mind that although over- or underrepresentation is proven, it is impossible to conclude what share of the attraction/repulsion is caused by what combination of nearby transport elements. In the case where multiple nearby elements increase attraction/repulsion, the results are affected in a more extreme fashion than would be expected when using the actual number of elements and actual observation area.

The following symbols will be used:

f	Error due to using expected values instead of actual values
n_{tr}	Number of transport elements
x	Number of overlapping areas $\in [1; n_{tr}]$
N_{act}	Actual number of land-use elements in whole area
N_{est}	Estimated number of land-use elements in whole area
M_x	Number of land-use elements in all areas where x observation areas overlap
A_{act}	Actual area
A_{est}	Estimated area
B_x	Area of all areas where x observation areas overlap
$r_x = \frac{N_x}{N_{act}}$	Ratio of elements in area x compared to total elements
$p_x = \frac{A_x}{A_{act}}$	Ratio of area in area x compared to total area

Suppose the study area is divided into different parts based on the number of overlapping observation areas. In other words, the surface where no overlapping areas overlap ($x = 1$) is the first area, where two observation areas overlap ($x = 2$) is the second area, etc. The number of actual land-use elements is the sum of the number of land-use elements in each of these areas separately.

$$N_{act} = \sum_{x=1}^{n_{tr}} M_x \quad 0.2$$

However, in the calculation as used in the data analysis, land-use elements within overlapping areas are counted multiple times (one time for each observation area):

$$N_{est} = \sum_{x=1}^{n_{tr}} x M_x = N_{act} + \sum_{x=1}^{n_{tr}} (x-1) M_x \quad 0.3$$

Similar relations can be set up for the area:

$$A_{act} = \sum_{x=1}^{n_{tr}} B_x \quad 0.4$$

$$A_{est} = \sum_{x=1}^{n_{tr}} x B_x = A_{act} + \sum_{x=1}^{n_{tr}} (x-1) B_x \quad 0.5$$

The error due to overlapping observation areas can be described as follows (this follows directly from the left hand term of the overrepresentation formula above):

$$f = \frac{f_N}{f_A} \quad 0.6$$

Where the two factors can be described as follows:

$$\begin{aligned} f_N &= \frac{N_{est}}{N_{act}} \\ &= \frac{N_{act} + \sum_{x=1}^{n_{tr}} (x-1) M_x}{N_{act}} \\ &= 1 + \frac{\sum_{x=1}^{n_{tr}} (x-1) M_x}{N_{act}} \\ &= 1 + \sum_{x=2}^{n_{tr}} (x-1) r_x \end{aligned} \quad 0.7$$

$$\begin{aligned} f_A &= \frac{A_{est}}{A_{act}} \\ &= \frac{A_{act} + \sum_{x=1}^{n_{tr}} (x-1) A_x}{A_{act}} \\ &= 1 + \frac{\sum_{x=1}^{n_{tr}} (x-1) A_x}{A_{act}} \\ &= 1 + \sum_{x=1}^{n_{tr}} (x-1) p_x \end{aligned} \quad 0.8$$

Substituting the above:

$$f = \frac{f_N}{f_A} = \frac{1 + \sum_{x=1}^{n_{tr}} (x-1) r_x}{1 + \sum_{x=1}^{n_{tr}} (x-1) p_x} \quad 0.9$$

In words: the error due to using estimated values of $n_{lu,observed}$ and A_{oa} instead of actual values depends on the ratios of the number of land-use elements in area x in combination with the ratios of the sizes of area x to the total observation area size. Three different situations can be described for any area x :

1. The over- or underrepresentation of the considered land-use element does not change in comparison to other x . In this case, the relative number of land-use elements found within x is the same as the relative area of x (compared to all elements and the whole area respectively), thus:

$$\begin{aligned} r_x &= p_x \\ f &= 1 \end{aligned}$$

2. The attraction of the considered area is higher than other x :

$$\begin{aligned} r_x &> p_x \\ f &> 1 \end{aligned}$$

In other words: the ratio of land-use elements within area x is larger than the area ratio of x . Additionally, since both r_x and p_x are multiplied by $(x - 1)$, the error increases for areas for larger x 's.

3. Similarly, the attraction is lower than other x :

$$\begin{aligned} r_x &< p_x \\ f &< 1 \end{aligned}$$

Concluding, the results only remain errorless when we assume that multiple nearby transport elements do not increase the over- or underrepresentation of land-use elements. However, differences are expected and therefore an error will occur in the results.

- Higher attraction when more elements are nearby, increase the overestimation of the estimated density. This overestimation could theoretically be infinite.
- Lower attraction when more elements are nearby, underestimates the estimated density. This underestimation could theoretically be infinite.

Therefore, the same results can be caused by multiple attraction configurations. It should not be forgotten that when over- or underrepresentation is found, is it without any doubt present in the map. It is, however, uncertain whether this is caused by attraction due to a nearby transport element independent of the *number* of nearby elements, or that a few nearby elements have no attraction and many nearby elements have some attraction. The results would be the same for both situations.

In order to fully avoid the problem, the number of land-use elements $n_{lu,observed}$ and the observation area size A_{oa} should not be estimated. To achieve this, land-use elements should never be 'counted' more than once (for computing algorithms, this would mean keeping track at which elements was already evaluated) and the determination of the area size should be based on the total map instead of each observation area separately. Only when these measures are done simultaneously, the error will be avoided. Only implementing one of these measures will make the error worse, since they cancel each other out in a part of the measurements.

Appendix H: Settings of other parameters in simulations

The random factor R was actually not included in the transition potential formula 6.1. The idea behind the random factor is to capture the unpredictable nature of human decision making. However, it appeared that the random factor sometimes had a surprisingly large effect on the simulation. In order to make all four simulations comparable, the random factor was turned off. However, it needs to be stressed that in applications of the model (instead of studies on the model itself, like this one), it should always be turned on. Please also refer to section 6.5.2.

The zoning plans include Natura 2000 data. The areas belonging to these protected natural areas were set to 'absolutely restricted' for both residential and commercial developments. This means that the value of $Z_{f,c}$ is set to 0. Consequently (see formula 6.1), $TP_{f,c}$ will be 0 as well. Thus, it is not possible that urban land-use will appear in the Natura 2000 areas. For forests and nature, zoning was set to 'stimulated', which entails a value for $Z_{f,c}$ of 1. For 'artificial green', it was set to 'unspecified'. Suitability maps were not included.

Zonal accessibility parameter values		
Minimal zonal accessibility		0.75
Sensitivity to costs (per activity)	Population	1
	Employment	0.9
	All other	0

Table 0.1 – Zonal accessibility parameter values (all simulations)

	Implicit accessibility of built-up areas	Implicit accessibility of non-built-up areas
Residential areas	1	0.3
Commercial and industry	1	0.3
Artificial green	1	1
Forests	1	1
Nature	1	1

Table 0.2 - Implicit accessibility parameter values (all simulations)

	Impassable
Other agriculture	No
Pastures	No
Residential areas	No
Commercial and industry	No
Artificial green	No
Forests	No
Nature	No
Wetlands	Yes
Fresh water	Yes
Marine water	Yes
Infrastructure	No
Mines	No
Port areas	No
Ports	Yes

Table 0.3 - Implicit accessibility parameter values (all simulations)

Appendix I: Adjusted local accessibility measure

The adjustments that were made to Metronamica's local accessibility component in order to dynamically categorise roads to flow is described here.

The local accessibility for each cell-link type combination is originally determined as follows. Only this part of the local accessibility model was changed; the subsequent parts that combine this into one local accessibility value for each cell ($LA_{f,c}$) will remain the same.

$$LA_{s,f,c} = \begin{cases} \frac{a_{s,f}}{D_{s,c} + a_{s,f}} & \text{if } a_{s,f} > 0 \\ 0 & \text{if } a_{s,f} = 0 \\ 1 - \frac{|a_{s,f}|}{D_{s,c} + |a_{s,f}|} & \text{otherwise} \end{cases} \quad 0.1$$

Where

Variables

$LA_{s,f,c}$ local accessibility of cell c to link type s

s link type

c cell

f land-use function of cell c

$D_{s,c}$ distance between c and nearest cell that is covered by link type s

Parameters

$a_{s,f}$ distance decay ('importance for land-use f to be close to link type s ')

The adjustment consists of a further classification of each link type s depending on its traffic characteristics. In the study case of the Netherlands, each link type s is further classified into three categories determined by the flow. The idea is to choose the distance decay parameter a depending on the flow level of the nearest link. Note that this does not mean the different flow levels are seen as separate link types.

Adjusting variables s and a and the local accessibility formula:

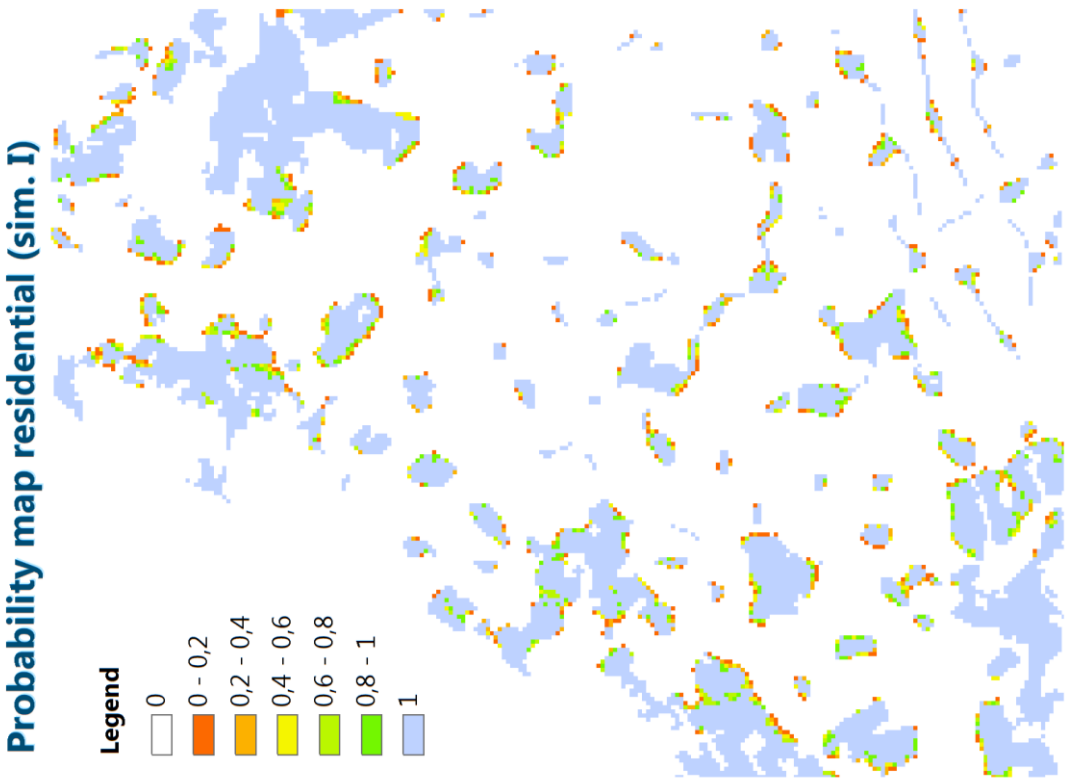
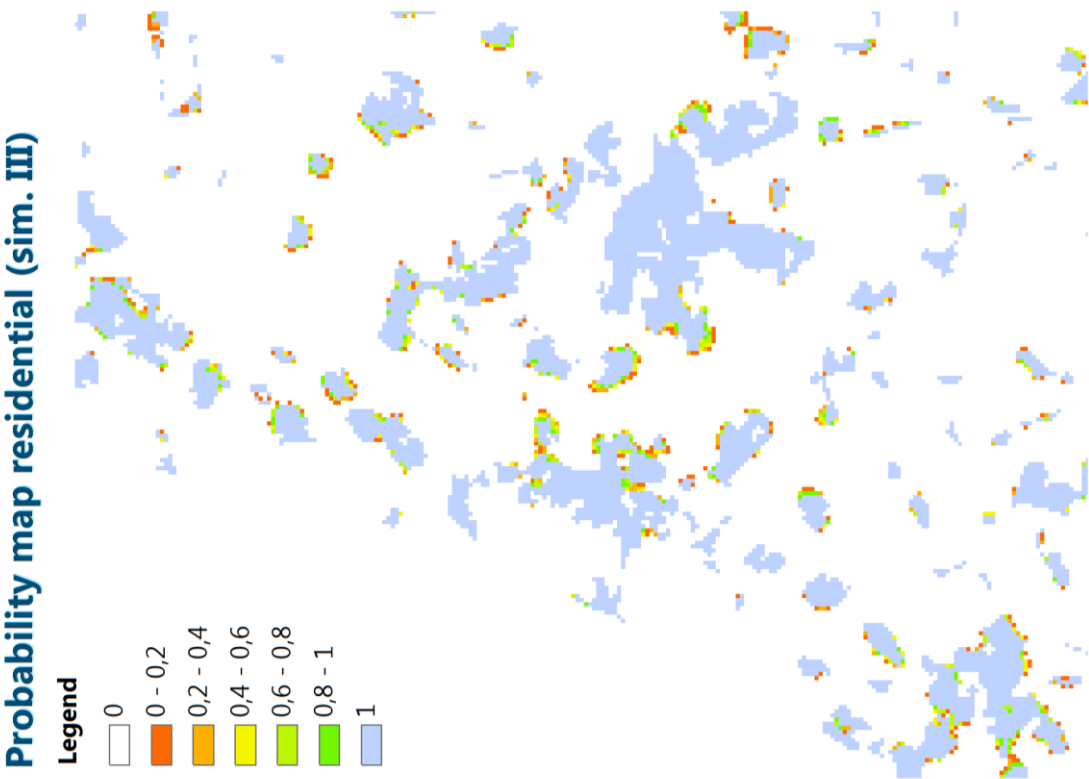
s_t link type s of category t

$a_{s_t,f}$ distance decay

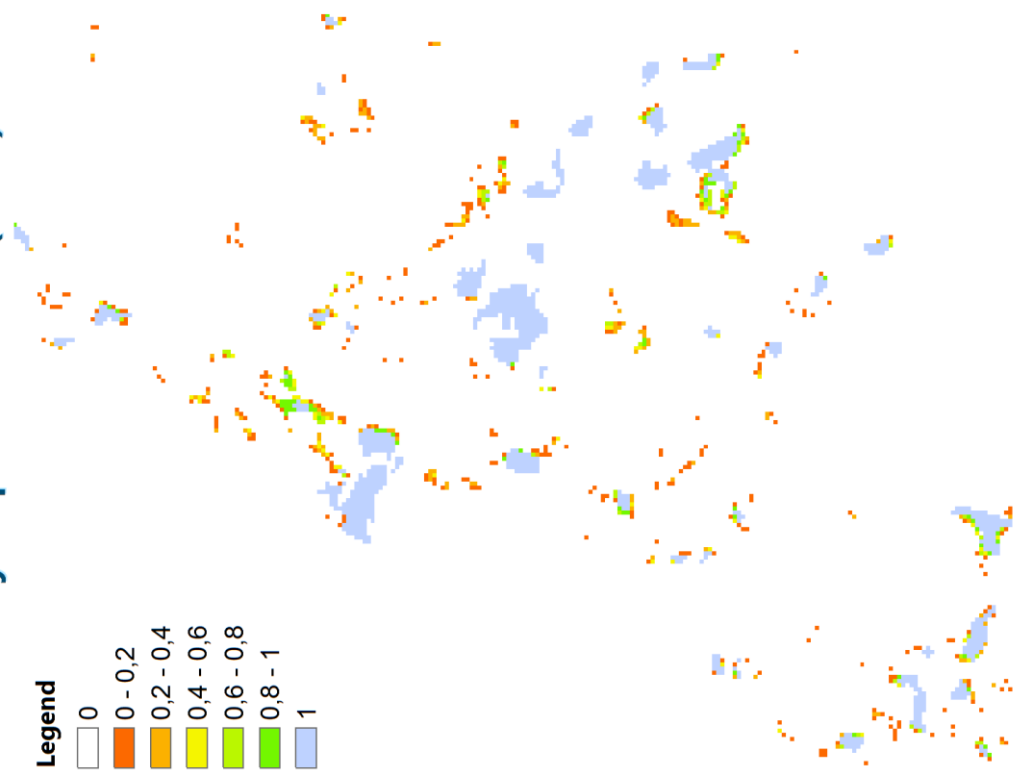
$$LA_{s,f,c} = \begin{cases} \frac{a_{s_t,f}}{D_{s,c} + a_{s_t,f}} & \text{if } a_{s_t,f} > 0 \\ 0 & \text{if } a_{s_t,f} = 0 \\ 1 - \frac{|a_{s_t,f}|}{D_{s,c} + |a_{s_t,f}|} & \text{otherwise} \end{cases} \quad 0.2$$

Appendix J: Probability maps

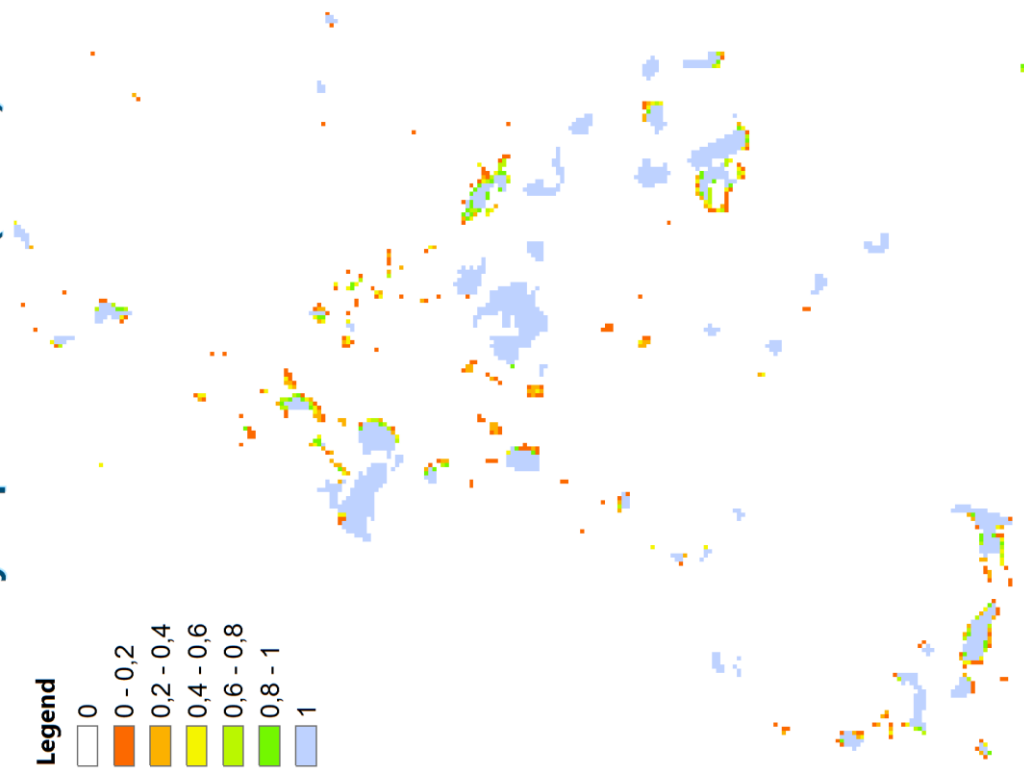
Experts from Randstad.



Probability map commercial (sim. I)

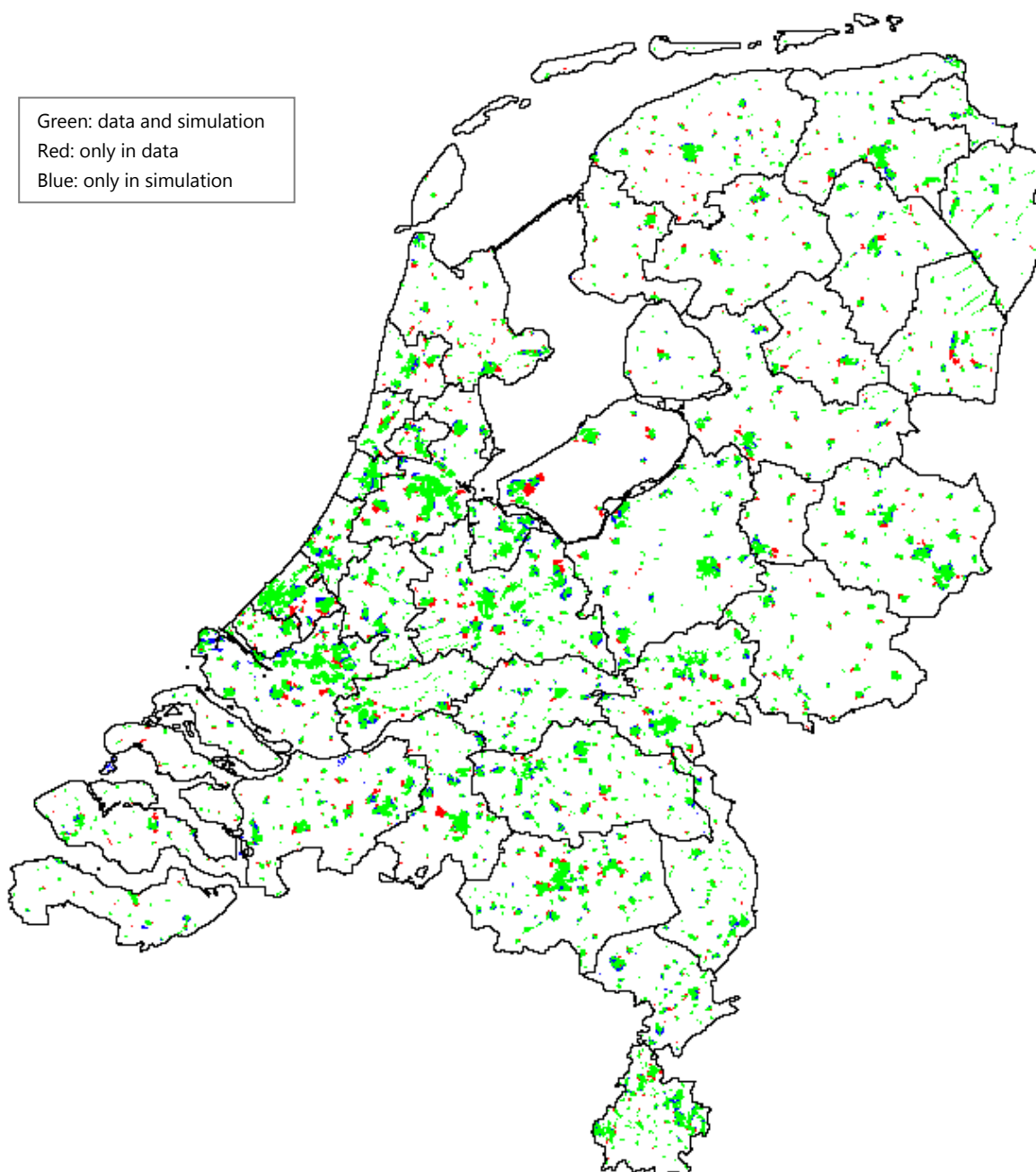


Probability map commercial (sim. III)

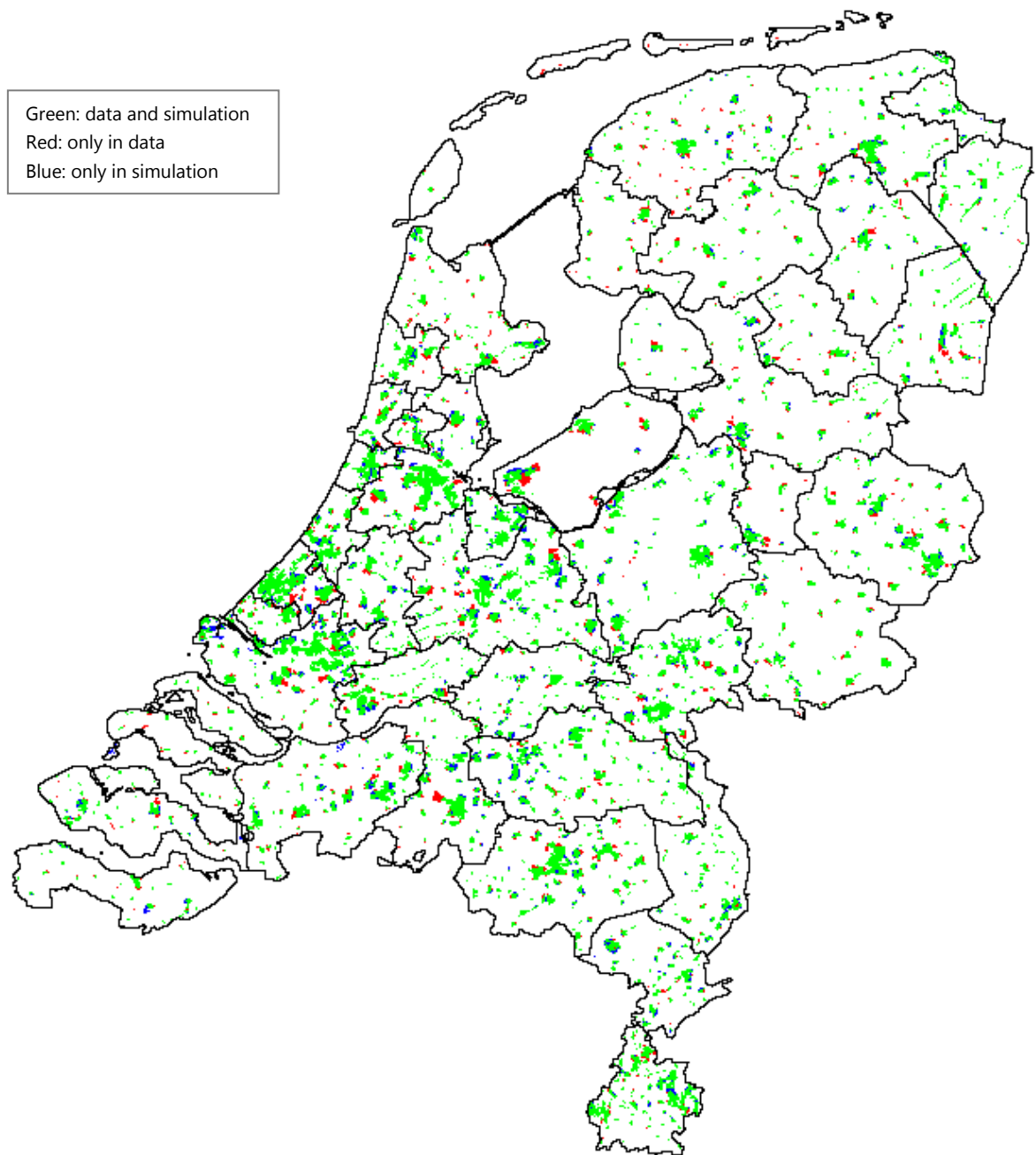


Appendix K: Enlarged maps

Data 2000 vs. simulation I 2000 (residential)



Data 2000 vs. simulation III 2000 (residential)



Data 2000 vs. simulation I 2000 (commercial)



Data 2000 vs. simulation III 2000 (commercial)

