Integration of present and future accessibility levels as spatially continuous utility factors in Land Use Scanner

Chris Jacobs-Crisioni

Vrije Universiteit, Amsterdam, 2011

COLOPHON

TITLE

Integration of present and future accessibility levels as spatially continuous utility factors in Land Use Scanner
Spinlab Research Memorandum SL-10

AUTHOR

Chris Jacobs-Crisioni, Spatial Information Laboratory (SPINlab), VU University Amsterdam.

CONTACT

VU University Amsterdam
Faculty of Economics and Business Administration
Department of Spatial Economics / Spatial Information Laboratory (SPINlab)
De Boelelaan 1105
1081 HV Amsterdam
the Netherlands
Phone: +31 20 5986125
Email: c.jacobs@vu.nl
Website: http://www.feweb.vu.nl/gis

Cover design: Irene Pleizier, SPINlab

This project was partially funded by and carried out within the Dutch National Research Programme 'Knowledge for Climate'.

CONTENTS

1 Introduction

In 2010 the Netherlands Environmental Assessment Agency (PBL) updated the so-called Land Use Scanner. In this update accessibility calculations have been integrated in the model. These can be used to provide more detailed, interval-specific utility factors for land use types that sustain human activities. Some features of the integrated accessibility calculations are noteworthy. First, the presented method uses a spatially asymmetric technique to sample and interpolate accessibility levels to a spatially continuous grid. Second, the presented method disaggregates regional employment and inhabitant projections to a spatially continuous grid. The integrated accessibility calculation method and its features will be presented in this paper. This report starts with an overview of Land Use Scanner, the role of accessibility in land use change and a review of what role accessibility played in previous versions of Land Use Scanner.

1.1 The Land Use Scanner model

Land Use Scanner is a doubly constrained logit model that allocates most probable land uses to a spatially continuous data grid (Hilferink and Rietveld, 1999). It uses Geographical Information Systems (GIS) methods and operations intensively. Initial versions of the model allocated shares of land uses to grid cells. More recent versions allocate discrete land uses (Koomen et al., 2008). The model is constrained by the amount of available grid cells in an allocation zone. The model is furthermore constrained by exogenously specified demands for land use types. These demands are derived from sectoral models like the `Pearl' model for housing demand and the `Bedrijfslocatiemonitor' for employment projections (see SPINlab research memorandum SL-05). The exogenous land use demands are specified in the model as regional land use claims. These are amounts of *area* necessary for a land use sector in a specific region. Notably, such land use claims can only be allocated to cells within the spatially discrete 'allocation regions' for which portions of these demands are specified.

Within the mentioned constraints, Land Use Scanner applies a logit model to allocate most likely land uses to grid cells. This logit model is based on differently weighed utility factors for a number of urban, commercial, industrial, natural and agricultural land use types. These utilities are derived from *suitability maps,* which in turn are based on "physical properties, operative policies and distance relations" (Loonen and Koomen, 2009; p. 15).Previously, static accessibility indicators have been used as suitability maps in Land Use Scanner applications. The work presented in this report deals with improving the consistency and validity of such suitability maps, by using present and future accessibility levels as a utility factor for land uses that sustain human activities. In the next section, accessibility is described as a driving force in urbanization processes.

1.2 The role of accessibility in land use change

From the work of Alonso (1964) follows that economic opportunities play an important role in location decisions and that such economic opportunities vary across space; according to Alonso, by variation in costs of access to the economic centre. We can expand Alonso's theory to take into account the complex spatial patterns of employment in modern polycentric urban landscapes, and state that *access to job-markets* is an important location factor for households. Let us expand

Alonso's theory even further and state that not just job-market access, but *the opportunity tointeract* is central in location decisions of economically driven actors, and thus, in economically motivated land use changes. In this light an indication of the amount of interaction opportunities ought to be central in explaining and predicting land use change.

There is overwhelming empirical evidence for such a central role of interaction potential in location decisions. In a seminal work Hansen demonstrates that placeswithbetter access to jobs, people or shops are more likely to be developed into residential areas (Hansen, 1959).Wegener and Fürst note that accessibility is an essential factor for retail, office and residential land uses(Wegener and Fürst, 1999).
Recent evidence from the Netherlands confirms the role of accessibility in urbanization processes. Priemus and Hoekstra have stressed the influence of interaction opportunities on location decisions of households and companies (Priemus and Hoekstra, 2009). Others highlight the role of easy access: proximity of transport-system entry points such as highway exits, train stations and airports (Atzema et al., 2009; De Graaff et al., 2007). Evidence that both interaction potential and infrastructure proximity are important for urbanization processes has recently been demonstrated for a number of cities (Borzacchiello et al., 2009; Borzacchiello et al., 2010).

Zondag and Pieters have done cross-sectional analyses on the influence of accessibility (using a logsum approach) on different types of households (Zondag and Pieters, 2005). They find that accessibility has an impact on the decisions of households to stay or move: households in more accessible areas are less likely to move. They furthermore find that traveltime to work is an important factor for location decisions, and that, even for individual households, the neighbourhood's interaction potential has a small yet significant role in location choices. De Bok and Van Oort have done a similar log-sum based cross-sectional analysis of how accessibility influences firm relocation choices. They find that, when relocating, all industry sectors have a preference for locations with good access to labour markets and the workforce. They furthermore find that the influence of the proximity of transport-mode accesspoints (highway ramps, stations) differs greatly between these sectors (De Bok and Van Oort, 2007).

1.3 Accessibility in previous Land Use Scanner configurations

Access to social and economic opportunities strongly influences land use change. In Land Use Scanner this access to social and economic opportunities has previously been proxied by indicators such as distance to a nearest highway exit, distance to a nearest train station, travel time to the nearest urban area or travel time to airports (seeTable 1). Ease of access is clearly accounted for by these indicators. However, these indicators are not well suited to describe the utility of interaction opportunities.

Table 1 - available access indicators used as utility factors in previous Land Use Scanner studies

Indicator	Type
Distance to nearest mainports in the Netherlands	Distance over road
Distance to nearest railway station in the Netherlands	Distance over road
Distance to nearest highway exit in the Netherlands	Distance over road
Distance to 100,000 inhabitants in the Netherlands	Distance over road
Distance to nearest urban area in the Netherlands	Distance over road
Traveltime to Schiphol (airport) and Rotterdam (port)	Traveltime over road
Distances to a number of economic and transport hubs in the EU	Euclidean distance
Distance to airports in the Netherlands	Partially Euclidean distance, partially distance over road

The limitations of the previously used access indicators will be demonstrated by one exemplary indicator. There are several reasons why `distance to a nearest large city' is not a well suited indicator for access to social and economic opportunities. First, this indicator underestimates the amount of economic opportunities for a location amidst multiple large cities. Next, the network distances used in 'distance to a nearest large city' are hardly inadequate to proxy of spatial separation, now that high-speed transport networks ever stronger differentiate the effort necessary to reach connected versus not-connected locations. Time or generalized costs are more adequate measures of spatial separation. Furthermore, the 'distance to a nearest large city' indicator does not distinguish in level of opportunities between differently sized cities. Lastly, such a static indicator does not take into account future infrastructural investments or land use changes.

Thus `distance to a nearest large city' is limited in indicating interaction opportunities. Other indicators that have previously been used in Land Use Scanner applications have similar limitations. These limitations have been overcome by implementing a more advanced interaction opportunity indicator.

1.4 This report

This report presents the method to calculate present and future accessibility levels that has been integrated in the Land Use Scanner framework. The implemented method is presented in section 0. The integrated accessibility measures are interpolated into spatially continuous factors. This method of spatial inference and tests on its reliability are presented in section 0. For the calculation of future accessibility, future spatial distributions of social and economic opportunities need to be estimated and the form of future transport networks needs to be known. This is elaborated upon section 0.

Some exemplary results of the method are presented in section 5. The data and software used in this study are presented in section 0. Some technical issues encountered in this study are elaborated upon in this section as well. Lastly, conclusions are drawn and further work is recommended in section 0.

2 The integrated accessibility calculation method

Several accessibility measures have been described in publications such as (Geurs and Ritsema van Eck, 2003; Rietveld, 1989). A common measure is applied in the presented work. The desired utility factors need to describe the spatial distribution of social and economic opportunities for specific land uses; i.e. how access to jobs, services or other people vary over space. Potential accessibility measures are adequate to indicate such spatial variance. Interaction opportunities are therefore calculated as potential accessibility measures as expressed in equation (1).

$$A(P)_{\,i,m,t} = \sum_{j=1}^{n} \frac{P_{j,t}}{f\left(c_{ij,m,t} + \ c_{jj,m}\right)} \tag{1}$$

In which $A(P)_{i,m,t}$ is calculated for zone i at time t, given connectedness by transport mode m with all zones j and opportunities P. This measure indicates the amount of opportunities P that can be reached from location i. The boundary between *can reach* and *cannot reach* is fuzzy: it is based on a distance decay function $f\left(c_{ij,m,t}\right)$, which in turn is based on lowest travel costs c between origin i and destination j at time t, given travel mode m. These travel costs are derived as travel times from a shortest path algorithm that is applied on transport-network data within Land Use Scanner. In the examples demonstrated in this study, accessibility measures are only calculated for passenger car transport.

The presented potential accessibility measure does not take into account that the opportunities at potential destinations might be non-replenishable, so that the actors that utilize these opportunities compete with each other. This non-competitive method is consistent with the overarching land use modeling framework. After all, the effect of competition for opportunities on regional growth is already accounted for in the models that supply the exogenous regional land use claims.

2.1 The implemented distance decay function

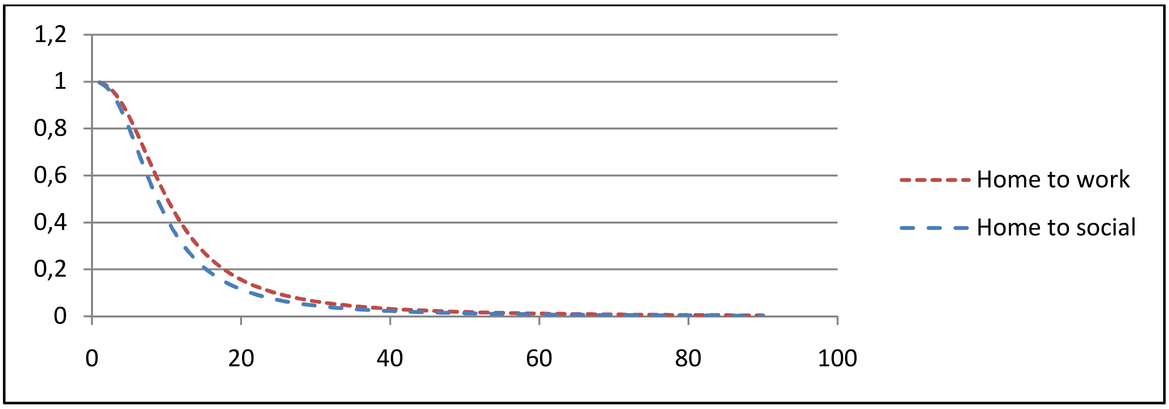
The results of the accessibility calculations depend on the formulation of the distance decay function $f(c_{ij,m,t})$. This is a function that relates the likelihood of an interaction to a measure of spatial separation. A log-logistic distance decay function is implemented. This form of function yields values between 0 and 1, and therefore has some particular benefits. Geurs and Ritsema van Eck (Geurs and Ritsema van Eck, 2003) mention the applicability of this function as a `fuzzy contour measure', in contrast with often used power and exponential functions. Furthermore, the applied log-logistic function yields much lower distance decay estimates at short distances and so reduces the systematic overestimation that exponential and power specifications risk with short distances. The implemented function is expressed in equation (2).

$$f\left(c_{ij,m,t}\right) = \left[1 + exp\left(a + b \ln c_{ij,m,t}\right)\right]^{-1} \tag{2}$$

In which the likelihood of an interaction $f(c_{ij,m,t})$ between locations *i* and *j* is based on travelcost *c* (given travelmode *m* at moment *t*) and opportunity and motive specific parameters *a* and *b*.

In a further section, exemplary social opportunity and employment opportunity indicators are presented. These are both based on the chance of an interaction occurring *from the home*. The resulting measures are therefore particularly valid for assessing the suitability of locations for residence. For the presented examples parameters have been used that are estimated for the aggregate Dutch population by Hilbers and Verroen (Hilbers and Verroen, 1993, p. 71); see Figure 1. For social opportunities, the parameters for trips with social motives are implemented, with *a* = -5.336 and *b* = 2.426. For employment opportunities, distance decay parameters for home-work trips are applied, with *a* = -5.691 and *b* = 2.463. Local numbers of houses are used as a proxy for social opportunities *P*, while for employment opportunities, local employment figures are implemented. On a sidenote, the 2010 Land Use Scanner configuration discerns a number of residential, commercial and industrial land use types. An interesting extension of this study might be the estimation of distance decay functions for each of these land use types.

Figure 1 - shape of distance decay functions for trips with social or home to work motives estimated by Hilbers and Verroen, 1993. The accessibility levels calculated with these functions can be used as utility factors for residential locations.



2.2 A spatially asymmetric sampling technique

A method of spatial inference is necessary to apply the resulting accessibility levels as utility factors in Land Use Scanner. In that land use model, both utility factors and resulting land uses are modeled on a spatially continuous plane. However, the accessibility calculation method presented in this report yields spatially discrete results. In the presented work a spatial interpolation technique, inverse distance weighing, is applied. The reasons to do so relate to reliability of the spatial inference method and prevention of land use allocation anomalies. For a detailed explanation and comparison of methods, see section 0.

The applied spatial interpolation method implicates specific demands for the spatial distribution of zones *i*. In this interpolation method the sources of spatial inference (the zones for which accessibility is calculated) exert a measure of influence on the inferred values with a weight that is inversely proportional to the distance between the source of inference (the centroid) and the point of inference (the point for which an accessibility value is inferred). From this method follows that more isolated points exert influence over a larger area, and thus, that a more equal dispersion of zones leads to a more equal dispersion of influence.

It seems plausible that the implemented interpolation method is most reliable when all sources of interpolation have an equal influence on the outcome. The presented work is therefore based on the assumption that the applied interpolation method is ideally based on equally dispersed locations. However, secondary data sources rarely have equal spatial dispersion. Such secondary data are implemented in this study to indicate the spatial distribution of opportunities, such as employment. There is thus a mismatch between the spatial distribution of zones to which secondary data sources are commonly aggregated and the *ideal* spatial distribution of discrete points for which accessibility is calculated.

To overcome a mismatch between the ideal spatial distribution of the to-be interpolated locations and the spatial distribution of secondary data zones, zones are applied asymmetrically. In other words, the number and the spatial distribution of i differs from j. Thus, for equally dispersed *sample* locations, access is measured to postcode-area based opportunities. Travelcosts are therefore calculated from the sample locations to the mentioned postcode areas.

2.3 Intrazonal distribution of opportunities

In the implemented method the intrazonal spatial distribution of opportunities is taken into account. The underlying assumption is that these opportunities are not concentrated in the geographical centers of their zones. A subsequent assumption is that, by taking the centre of a zone as the average point of destination, the method will underestimate the costs needed to reach the opportunities in that zone. These shortcomings of aggregated data might lead to a bias, in which the opportunities offered by larger zones are overestimated. To overcome this issue a solution to estimate intrazonal distances is adopted; see (Horner and Murray, 2002). This method is expressed in equation (3).

$$c_{jj,m} = \frac{\sqrt{\frac{A_j/\pi}{2}}}{s_m} \tag{3}$$

In which the internal impedance $c_{jj,m,t}$ for zone j, with area A, for mode m is defined, given an estimated friction s_m.In the implemented method the friction is solely based on travel speeds. As a rough estimate of the average speed on local streets a constant intra-zonal travel speed (s_m) of 40 km/h is used.

3 Spatial inference of accessibility levels

In the presented work accessibility levels are calculated for a number of discrete locations. However, the utility factors used in Land Use Scanner are derived from continuous data. Some form of spatial inference is therefore necessary to apply the implemented accessibility levels on a spatially continuous plane. Commonly, accessibility estimates are calculated for a discrete point in a zone (usually a centroid); all space in that zone is then inferred to have that location's level of accessibility. For examples, see (Geurs and Ritsema van Eck, 2003; Hilbers and Verroen, 1993). However, for two reasons the common zonal approach is unsatisfactory for the presented application.

First, a zonal inference approach for calculating the desired utility factors might lead to `border effects' in the (spatially continuous) land use allocation. These border effects occur most notably when inferred utility factors (such as accessibility) differ strongly between two bordering zones. Such differences can force the land use model to allocate land uses along often arbitrarily defined zonal borders, resulting in an unwanted influence of arbitrarily defined borders on modeled land use patterns.
Second, for computational reasons accessibility can only be calculated for zones on a fairly high level of spatial aggregation. It is likely that at increasing distances between the sources of inference, a point between neighbouring zone centroids is as accessible as the average accessibility of these sources. If so, a zonal approach for inferring accessibility levels is likely to be erroneous in the border areas of the zones. It is expected that an interpolation method reduces the errors between the sources of inference.

3.1 The implemented method of spatial inference

The accessibility values that result from equation (1)are interpolated to a spatially continuous grid by means of an inverse distance weighted method as expressed in equation (4).

$$A(P)_{g,m,t} = \left. \sum_{i=1,d_{gi}<5250}^{n} \left(\frac{A(P)_{i,m,t}}{\left(d_{ig}\right)^2} \right) \middle/ \sum_{i=1,\, d_{gi}<5250}^{n} \left(\frac{1}{\left(d_{ig}\right)^2} \right) \right. \tag{4}$$

In which the accessibility value *A(P)* for gridcell *g* is based on the Euclidean distance d_{ig} between zonal centroid *i* and grid cell *g.*
It is not plausible that accessibility levels from farther away are more related to the accessibility of zone *i* than the accessibility levels of direct neighbours. The centroids of zone *I* are approximately 6,500 meters apart. Therefore the influence of *i* on *g* is only calculated when $d_{gi} < 5{,}250$ meters. Thus, for any point in space only directly neighbouring zones are included in the interpolation of values.

3.2 Test of the implemented spatial inference method

It is expected that the interpolated accessibility estimates implemented in this study are more accurate approximations of accessibility at a particular point in space than zonal accessibility estimates. The following hypotheses are tested:

1) a larger degree of the variance in real accessibility (an estimate calculated for an exact location) can be explained with interpolated accessibility estimates; and
2) when comparing the performance of the interpolated and the zonal models, there is less systematic influence of the distance between the source of accessibility estimates and the point of interest in the error component of the interpolated estimates.

To test these hypotheses 1,500 random sample points s have been generated. Subsequently, directly calculated accessibility values as well as accessibility levels derived from two spatial inference methods are attributed to these points. First, for each of the points s the direct level of accessibility (the 'real' accessibility levels $RealAccess_s$). Next, accessibility values $InferredAccess(M)_s$,derived from the two spatial inference methods M (i.e. an interpolation and a zonal inference method) are assigned to the points s. These inferred values are both derived from the accessibility calculations that have been implemented in the presented work , which are calculated for the centroids of approximately 1,500 equally dispersed hexagonal zones i. The zonal inference values are derived from the zones I in which the sample points s lie. The interpolation values are derived from inverse distance weighing of the zonal accessibility values. On a sidenote, all accessibility values have been derived from the same network and the same spatial distribution of opportunities.
Errors for both zonal and interpolated inference methods have subsequently been calculated as expressed in equation (5).

$$InferenceError(M)_s = \left(\sqrt{(InferredAccess(M)_s - RealAccess_s)^2}\right) \Big/ RealAccess_s \tag{5}$$

The performance of inference methods is compared with a t-test. The deviation of values of $InferenceError(M)_s$is tested (see Table 2). The conducted test shows that with 95% certainty the errors of the interpolation method are smaller than the errors of the zonal method. Conclusively, for the presented application the interpolation method is slightly more reliable than zone-based spatial inference.

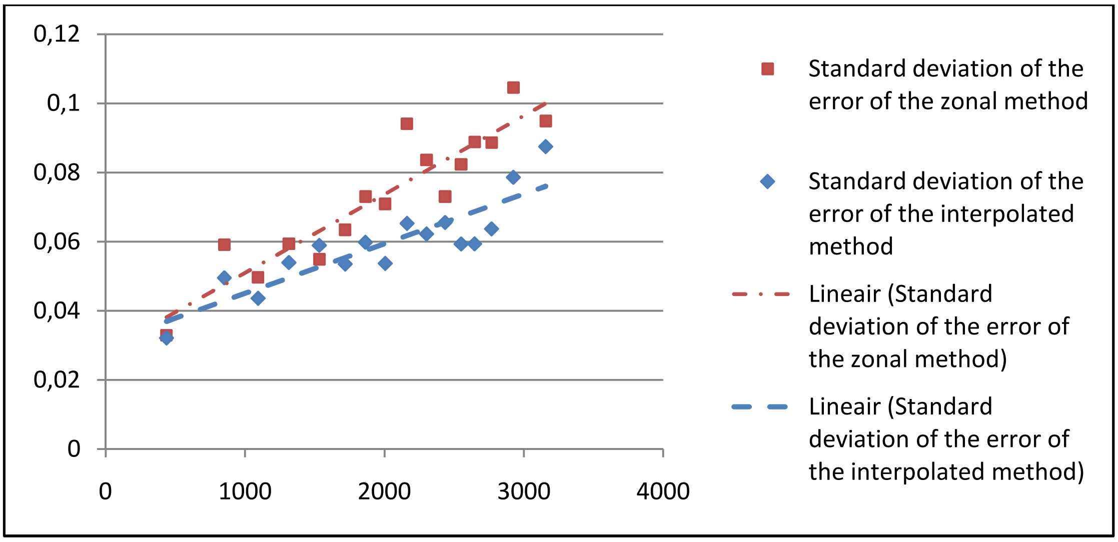
Table 2 - t-test results of spatial inference methods; test value = 0; n = 1500

Inference error of method	**Significance (2-tailed)**	**95% confidence interval (lower, mean, upper bound)**		
Interpolation method	0.000	0.042	0.044	0.046
Zonal method	0.000	0.054	0.057	0.059

To find if the interpolation method is systematically more reliable, more tests on the compared inference methods are necessary. It is expected that spatial interpolation performs better at larger distances from the sources of inference compared to the zonal inference method. Nevertheless, any method of spatial inference is presumably more likely to be erroneous when distance from the source of inference increases. Therefore the performance of inference methods over distance is compared. To do so, Euclidean distances d_{is} between sample point s and the nearest source of inference I have been calculated and classified into 15 classes with equal amounts of observations.

Figure 2 presents the standard deviations of errors of the two inference methods at average distance values of the 15 classes of d_{is}. At increasing distances d_{is} no systematic under- or overestimation of accessibility occurs. However, as the trend lines in Figure 2 suggest, the errors of the estimations deviate more when d_{is} increases. With Pearson correlation coefficients of respectively 0.880 and 0.924, deviations in the errors of the interpolation and zonal inference methods are both highly correlated with distance to the source of inference. It can be concluded that indeed, regardless of the method of inference, the error of spatial inference increases when the distance to the source of inference increases.
However, the trend line that indicates the standard deviation of errors of the zonal inference method increases more when d_{is} increases than the trend line of the interpolation method. This suggests that the interpolation method is better at reducing systematic errors of spatial inference that are conjoined with distance to the source of inference. It can thus be concluded that, for the application at hand, the interpolation method is systematically more reliable than the zonal inference method.

Figure 2 -standard deviations of differences between inferred and real accessibility estimates at the average equal-n classes of distances from a random point to the nearest source of spatial inference.



4 Calculating future accessibility levels

In the 2010 version of the Land Use Scanner, land use change is simulated at a number of intervals. At every interval, land use allocation is based on suitability maps that are derived from the allocation results of the previous interval. At each interval, the presented accessibility indicators are updated as well. For these indicators the values of P (i.e., housing or employment) and/or c (travelcost) might change. The methods to estimate changes in P or c at all intervals are elaborated upon in the next sections.

4.1 Disaggregating projections of future opportunities

For a proper estimation of accessibility levels, the spatial distribution of opportunities needs to be known at a fairly low level of spatial aggregation. Land Use Scanner only dedicates land use types to grid cells. The projected amounts of inhabitants and employment on which land use claims are based are only available at a regional level. Therefore, opportunities used in the accessibility calculations have to be derived by spatial disaggregation of regional projections.

A simple method is applied to disaggregate regional amounts of houses and employment. The amount of opportunities P per grid cell g (and land use lu) at time t in zone i is calculated as in equation (6).

$$DISAGGREGATE(P)_{lu,g,t,r} = \left(\frac{PROJECTED(P)_{lu,t,r}}{count\left(G_{lu,t,r}\right)}\right) \tag{6}$$

In which $DISAGGREGATE(P)_{lu,g,t,r}$ being the amount of opportunities P per gridcell g for land use lu at time t in region r; and $G_{lu,t,r}$ being gridcells G with allocated land use lu at period t in region r. The$PROJECTED(P)_{lu,t,r}$ is the projected amount of opportunities P for a land use lu at period t in region r.

This method implies the assumption that all projected amounts of houses and employment are allocated by the land use model; it furthermore implies the assumption that densities of population or employment have no spatial variance within the allocation zone. Especially the latter seems implausible. Therefore the next section presents tests of the performance of the presented disaggregation method as well as explorations of the variance of errors.

4.2 Test of the disaggregation method on current housing

The amount of disaggregated housing according to the integrated model $DISAGGREGATE(HOUSES)_{2006,z}$ is compared with observed distributions of $HOUSES_{2006,z}$ for a number of spatial aggregations z. The observed distributions of houses are derived from a dataset that contains observations by Netherlands Statistics (CBS) edited by VROM, RIGO and PBL. Model errors $ModelError_{houses,2006,z}$ are derived from differences between disaggregated and observed housing numbers as expressed in equation (7).

$$ModelError_{houses,2006,z} = \sqrt{\left(DISAGGREGATE(HOUSES)_{2006,z} - HOUSES_{2006,z}\right)^2} \tag{7}$$

Based on these model errors, the total model error $TotalModelErrorZ_{houses,2006}$ for spatial aggregations z is calculated as shown in equation (8).

$$TotalModelErrorZ_{houses,2006} = \sum_{z=1}^{n} ModelError_{houses,2006,z} \Big/ \sum_{z=1}^{n} HOUSES_{2006,z} \tag{8}$$

The results of *TotalModelErrorZ* are shown in Table 3. The housing disaggregation method performs better at higher levels of aggregation. The disaggregation method seems to perform rather well even with the relatively small '4-digit' postcode areas used in the accessibility calculations. On a sidenote, a problem with data aggregation becomes apparent when comparing total numbers of observed housing at different levels of spatial aggregation. These problems are described in section 6.3.

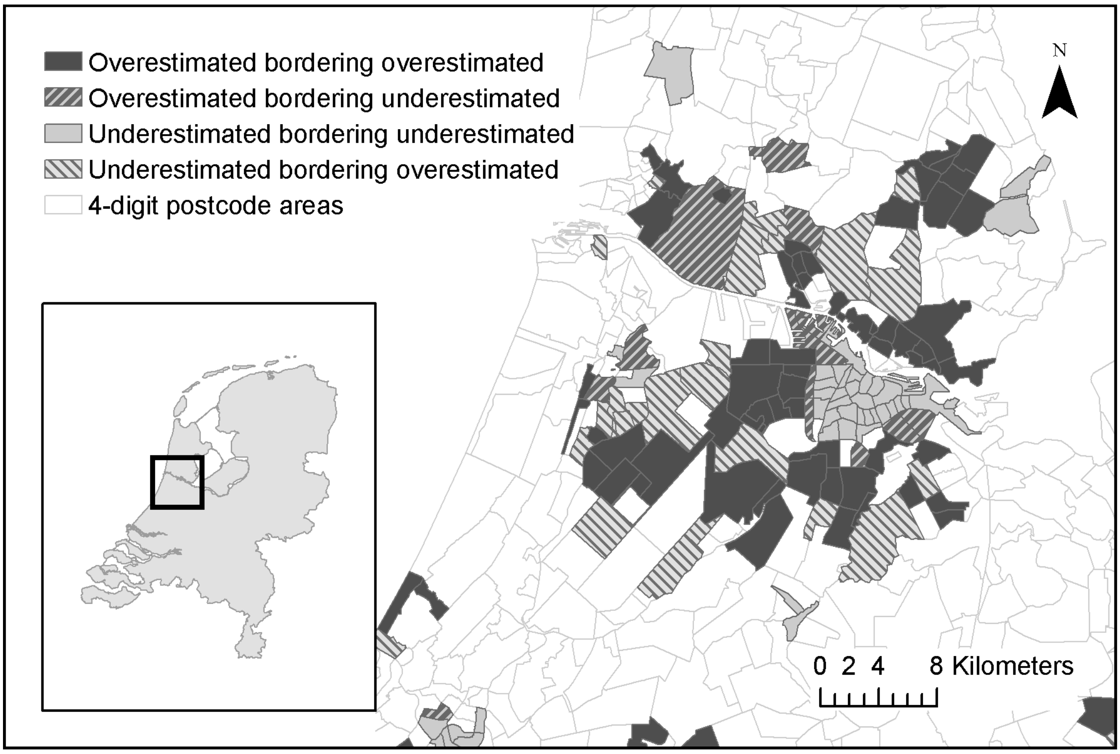
Table 3 -model performance at different aggregation levels

	Grid (n = 407,177)	**4-digit postcode area (n = 4,023)**	**Communities (n = 472)**	**COROP+ (n = 53)**
Number of observed houses (edited from CBS, 2006 by VROM, PBL)	7,087,053	7,079,975	7,085,792	7,085,792
Sum of absolute difference between estimated and observed	7,121,124	1,978,228	995,871	398,020
Error of housing estimation method	1.005	0.279	0.141	0.056

Possibly systematic factors in errors of the disaggregation method are explored. First, the expectation that land suitability affects housing densities is tested. Pearson correlation tests have been applied to find a relationship between residuals of the model (see equation (7)) and suitability values for dense residential areas as calculated in the `Leefomgevingsbalans 2010' configuration of Land Use Scanner. These correlation tests have been applied on the data at two levels of spatial resolution. On grid level, residuals and suitability levels have been compared without any need for aggregation; on 4-digit postcode level the residuals have been compared with *mean* suitabilities per postcode area. At both spatial scales there is little correlation between land use suitability and model residuals. At grid level, the Pearson correlation between model residuals and land use suitability is 0.078; at 4-digit postcode level, correlation is -0.16. It can be concluded that a straightforward inclusion of residential suitability factors in the disaggregation model is not enough to reduce the disaggregation model error.

Next to testing the residuals, an exploratory spatial data analysis of model residuals has been performed to find spatial patterns in the model error. The data is slightly clustered, with a Moran's I of 0.2511 (where 0 = no clustering and 1 or -1 = very clustered). A Local Indicator of Spatial Autocorrelation (LISA) analysis of the data (Anselin, 1995) shows some patterns in these clusters (see Figure 3).

Figure 3 - results of LISA cluster analysis of housing disaggregation model residuals in Amsterdam area.



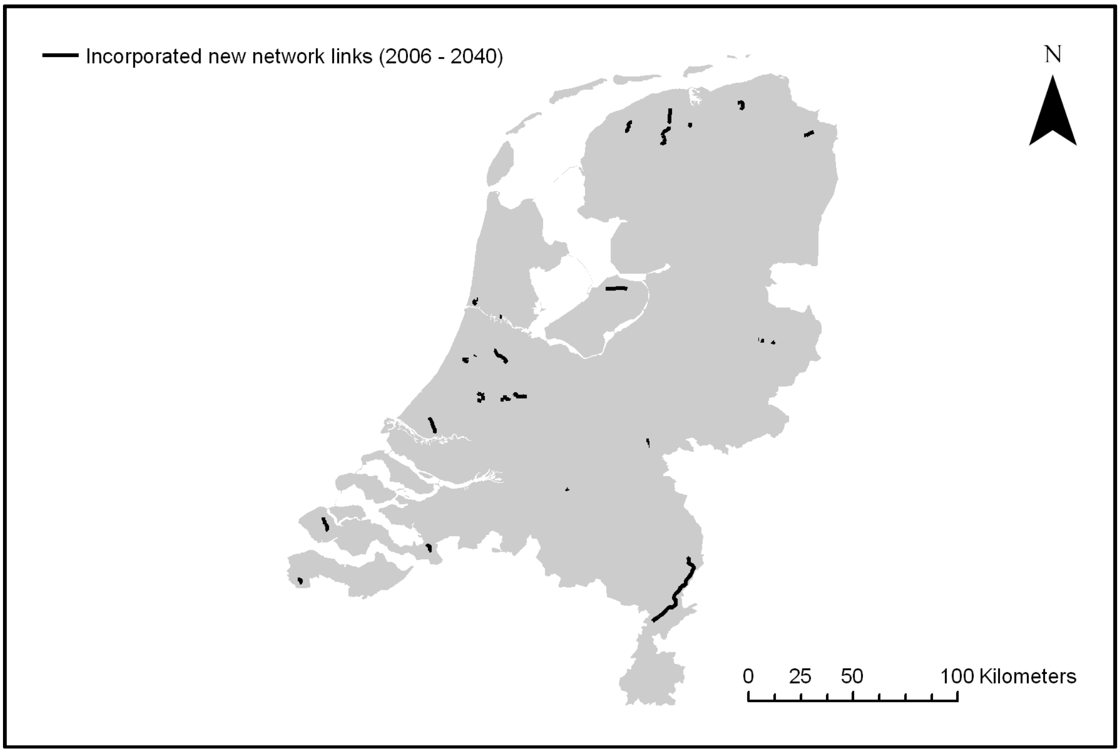
Notably, the central Randstad conurbation shows a large checkerboard of clustered areas where the housing disaggregation model underestimates numbers of houses in the city centres (e.g. in the centres of Amsterdam, Utrecht, Rotterdam and the Hague). The model furthermore overestimates numbers of houses in suburban areas (notably in city peripheries and in 'het Gooi', a wealthy suburban region near Utrecht).

Conclusively, the model underestimates housing densities in historical city centres, and overestimates housing in suburban areas. The suitability factors that are currently used in Land Use Scanner are not related to spatial variations in the model errors. A disaggregation model which takes historical housing densities into account should be able to solve this. Such a model might apply a nested structure in which 1) existing residential areas assume historical housing densities if no land use changes occur locally, while 2) the densities of new or changed residential areas are based on the modeled type of residential area and perhaps a proxy of the local utility for residential land.

4.3 Applying future changes in the road network

Next to social or economic opportunities, costs for traversing transportation networks are likely to change over time. Both congestion and infrastructural investments might affect future traveltimes and transport costs. In the presented methodology congestion effects are not yet implemented. However, infrastructure investments with a likely super-local influence have been incorporated as new network links. In the presented method, these new links are available for shortest path calculation after the planned time of opening. These network links have mostly been derived from a dataset of spatial planning decisions of national, regional and local administrations. Figure 4shows the incorporated network links.

Figure 4 - incorporated future links in the network.



5 Accessibility levels as utility factors

This chapter presents some thoughts on how to apply the presented accessibility levels as utility factors in Land Use Scanner. It furthermore presents some resulting accessibility maps and demonstrates how future network links affect local access to current employment opportunities.

5.1 Applying accessibility levels as utility factors in Land Use Scanner

Equation (1) results in absolute values bearing little meaning. To interpret and weigh the influence of accessibility on land use change these results ought to be rescaled into more meaningful values. Therefore the results of equation (1) are rescaled as shown in equation (9).

$$RA(P)_{i,m,t} = \frac{A(P)_{i,m,t}}{\max\ A(P)_{z,m,t}} \tag{9}$$

In which relative accessibility levels are calculated as RA (given opportunities P) for zones i in allocation region z, given transport mode m and interval t. Some relevant choices are implicit in this rescaling operation. By choosing to rescale to interval-specific maximum values, a *competitive* or *zero-sum* model of accessibility effects is applied (Rietveld, 1989). The potential effects of accessibility growth on regional economic structures are not taken into account. This is consistent with the current Land Use Scanner framework in which fixed amounts of land use are allocated to regions, without feedbacks on the regional models that project future land demands.
In Land Use Scanner, accessibility is thus used to simulate the intraregional competition of locations for land uses. This is reflected in the choice to rescale to the maximum values of allocation regions z. This rescaling stresses intrazonal variation of accessibility. However, the resulting spatial distribution of relative accessibility levels is prone to overstress the intraregional variation of accessibility. In peripheral regions this might lead to overestimation of the effect of relatively small differences in accessibility on land use development. Furthermore, rescaling at regional level introduces border effects in the calculated utility factors that are avoided when rescaling to one maximum value in the area of interest.

5.2 Accessibility maps rescaled to national and regional maxima

Maps that indicate the spatial variation of access to social and employment opportunities in 2006 in the Netherlands are depicted in Figure 5andFigure 6. These are based on observed land use, and observed regional housing and employment numbers that have been disaggregated to gridcell level. Figure 5shows the spatial variation of these accessibility measures at a national scale. Subsequently, maps of access to employment opportunities rescaled to regional maxima are demonstrated in Figure 6.

Figure 5 - access to social opportunities (top) and access to employment opportunities (bottom) in 2006 in the Netherlands as calculated by the presented method.

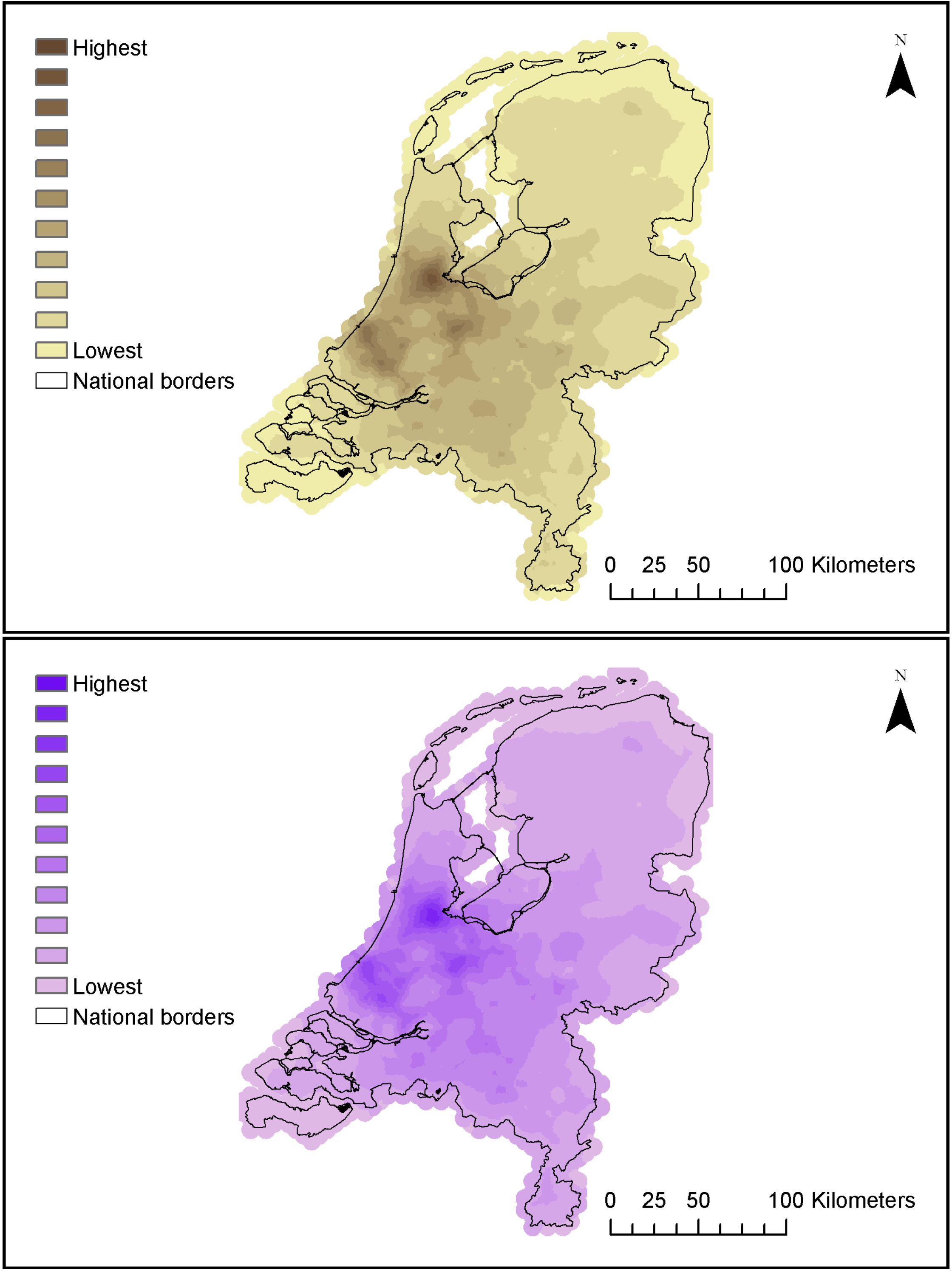
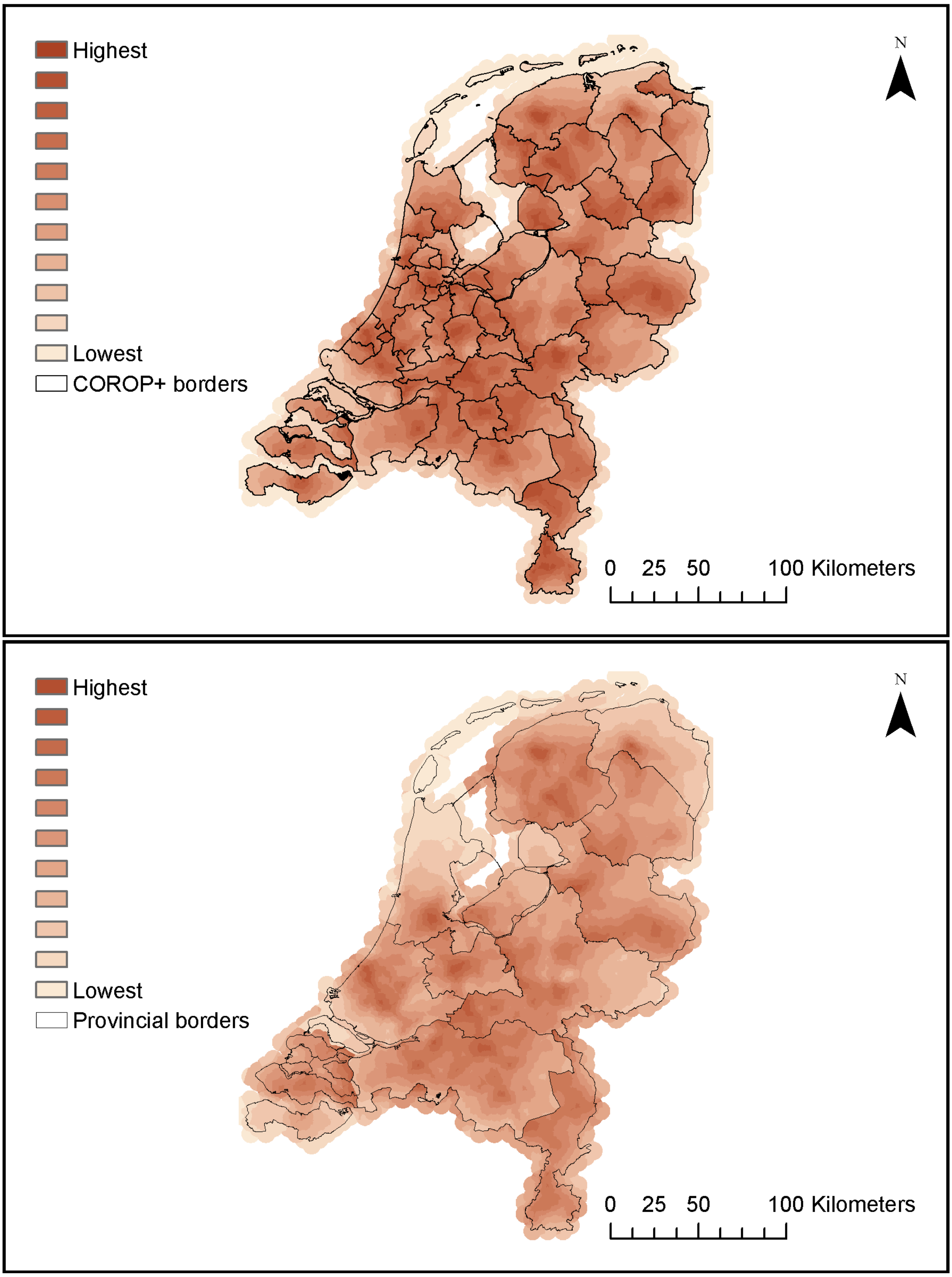


Figure 6 - Access to employment opportunities in 2006 in the Netherlands rescaled to zonal maximum values for different zonal divisions. In this case provincial (top) and regional (COROP+, bottom) maximum values have been used. In Land Use Scanner practice, the chosen regional level of land use allocation can differ per application.



5.3 Accessibility effects of future links

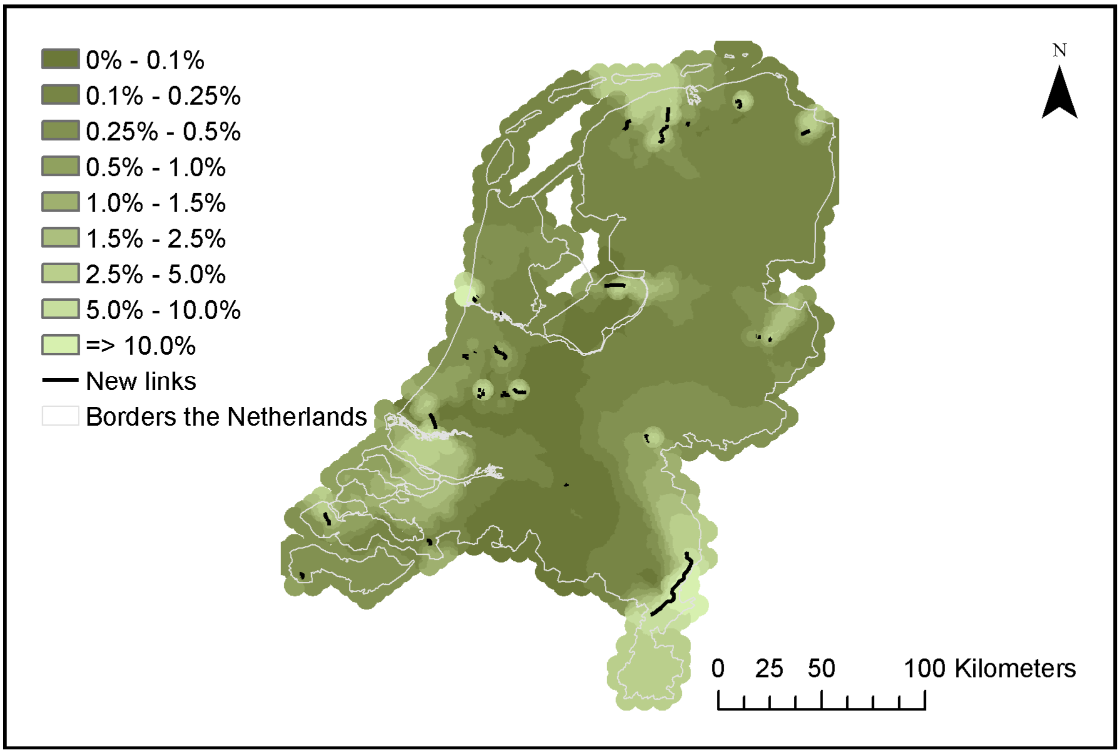
New network links, policy changes, capacity increases and other infrastructural investments will reduce the (free-flow) travel costs c for specific relations ij. Such changes are accounted for by using time-specific versions of the network from which travel costs are derived. However, only new links in the network are accounted for. The effects of congestion, capacity increases and pricing policies are excluded.

To visualize the effects of future links on accessibility levels, such accessibility levels have been calculated with both the current (2006) and future (2040) road network. These accessibility levels are based on observed numbers of jobs in 2006 derived from the `LISA' dataset (LISA, 2006). The resulting accessibility growth between 2006 and 2040 $\Delta A(JOBS)_i$ is then calculated as shown in equation (10).

$$\Delta A(JOBS)_i = \left. \sum_{j=1}^{n} \frac{JOBS\_2006_j}{f\left(c_{ij(2040)} + c_{jj}\right)} \middle/ \sum_{j=1}^{n} \frac{JOBS\_2006_j}{f\left(c_{ij(2006)} + c_{jj}\right)} \right. \qquad (10)$$

From these calculations follow percentual changes of accessibility, of which the interpolated form is depicted in
Figure 7. Most notably, this figure indicates that new network links between 2006 and 2040 strongly affect the Dutch periphery.

Figure 7 - percentual growth in access to jobs (level 2006) by passenger car bylinks added to the network between 2006 and 2040.Congestion effects and network capacity increases are not included.



6 Data and implementation issues

In this section the software and data used in the implemented method is presented. Furthermore, some general implementation issues are described.

6.1 Data used in the implemented method

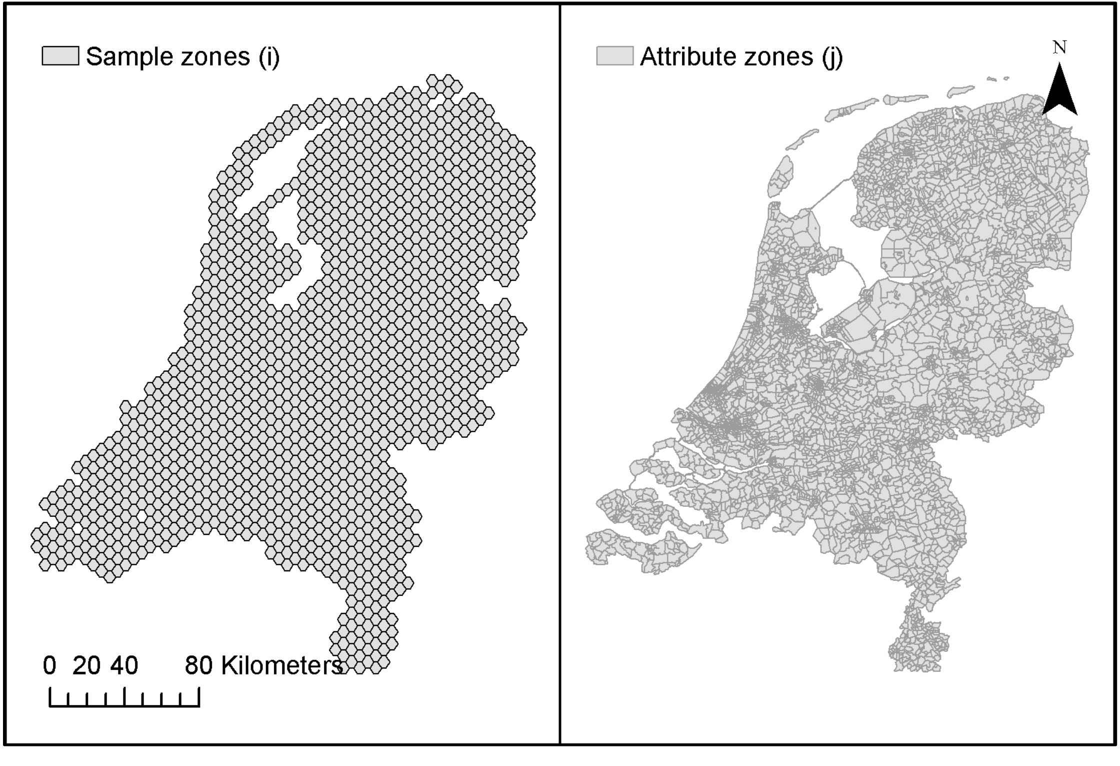
A number of secondary datasets are used in this analysis. Some key properties of these datasets are presented in Table 4.

Table 4 - key properties of the datasets used in the calculations. Size of the zonal data is in square kilometers of area, size of the streets data is length in meters.

Data	Year	Number of features	Minimum size	Mean size	Maximum size
Hexagonal zones	2010	1,504	28.17 sq. km.	28.17 sq. km.	28.17 sq. km.
4-digit postcode zones	2006	4,023	0.02 sq. km.	8.66 sq. km.	139.64 sq. km.
Navteq streets	2007	1,274,189	2.00 m.	122.19 m.	154,248.11 m.

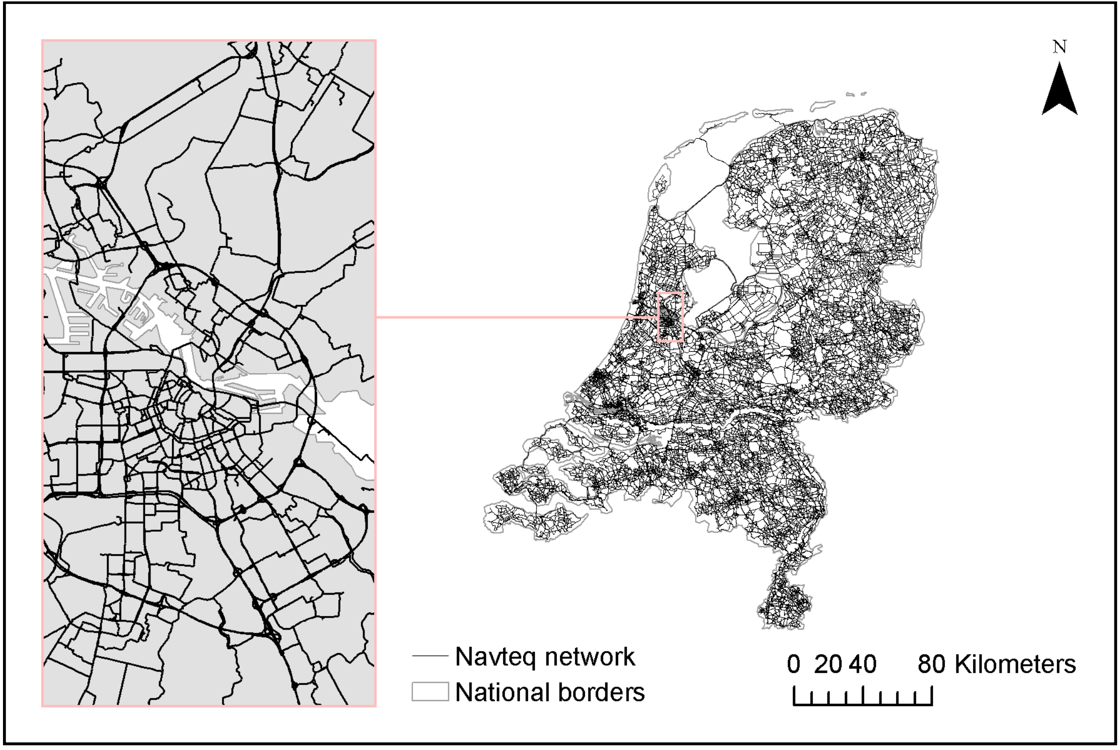
The accessibility values in zones *i* are calculated for the centroids of 1,504 hexagonally shaped sample zones. These hexagonal zones have beforehand been generated by means of a Vba script in ArcGIS. The opportunities in zones *j* are distributed over 4,023 centroids derived from so-called 4-digit postcode areas, for which the necessary secondary data is available. The zones used in the calculations are depicted in Figure 8.

Figure 8 - zones i (for which accessibility levels are sampled) and zones j (containing opportunity attributes) used in the presented method



The travel costs *cij* are derived from a road transport network composed by Navteq (2007). This network comprises all types of roads and streets in the Netherlands and includes ferry services. Travel time is applied as a proxy for travel costs. Only a portion (~255,000 of ~1,200,000) of the links in the original Navteq dataset is used; see Figure 9 for an impression of the implemented network, and see section 6.5 for details.

Figure 9 - the network used in the accessibility calculations



Before the used network is queried, the zones *i* and *j* are connected to the network. A common connection method is applied, which connects zones to the (in Euclidean distance) nearest link of the network. The connect operation is not allowed to connect to motorways, elevated sections of road or roads with a maximum speed of over 80 km/h. Furthermore, a length-related impedance factor is added to the new connector links that are created in the connect operation. The implemented impedance factor is based on a roughly estimated average inner-city travel speed of 40 km/h.

6.2 Source of future network links

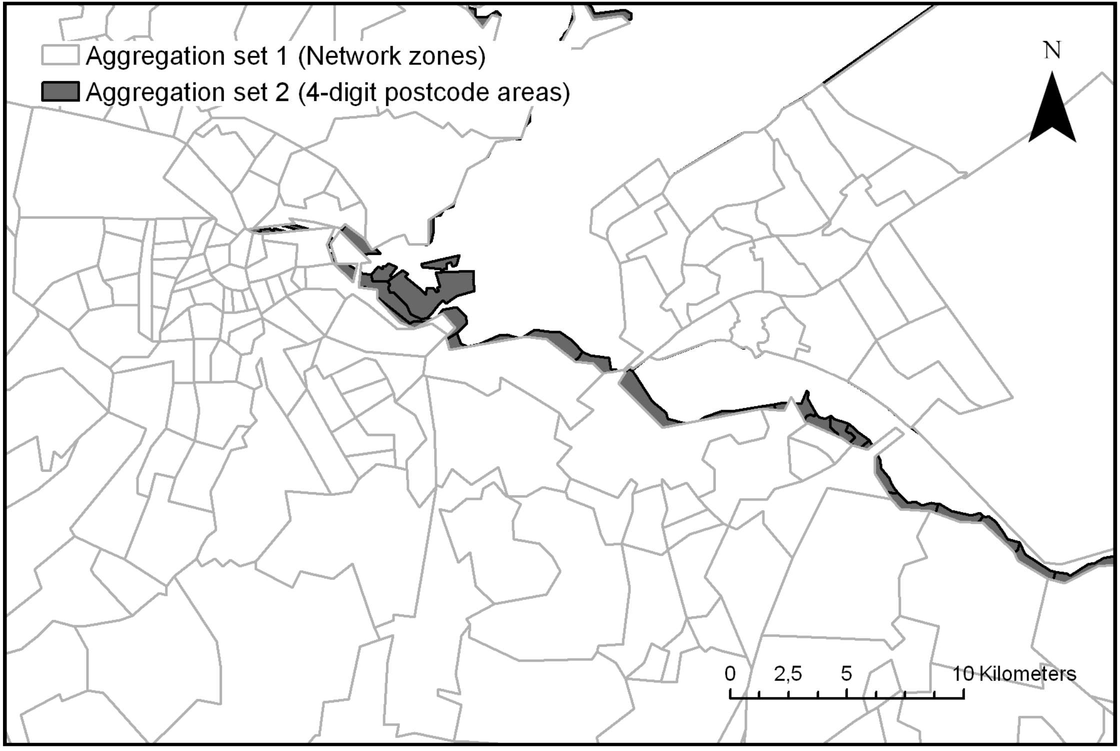
Future network links are derived from the so-called ‘new map of the Netherlands’ (NIROV, 2010). This map depicts spatial plans from spatial planning and infrastructure management authorities in the Netherlands. It includes an indication of the certainty that the depicted plan will be realized, and the planned moment of realization. Only planned road extensions that will be applied with high certainty have been incorporated. Furthermore, only links that appeared to serve super-local purposes have been incorporated. Lastly, the planned moments of realization have been used as moments of opening.

6.3 Data issues

The housing disaggregation test results unveil some problems with housing data, as well as problems with different zonal division datasets used in Land Use Scanner. First, a comparison of sources demonstrates that available datasets do not consent on the observed amount of houses in the Netherlands. Netherlands Statistics claims that the Netherlands contained approximately 6.9 million houses in 2006 (Statline, 2010); the dataset used in the model tests claims approximately 7.1 million houses in 2006 (CBS, VROM, RIGO, PBL; 2010). These differences influence the model reliability tests. These differences are presumably due to different techniques, criteria or moments of measurement and do not have a drastic effect on spatial distributions; if this assumption holds, the housing figures used in the accessibility calculations are reliable.

Second, geographical mismatches between different zonal divisions in Land Use Scanner have become apparent. In the housing disaggregation tests, housing numbers have first been disaggregated to grid level, and subsequently aggregated to a number of zonal division sets. The total number of aggregated houses then differs between the differently aggregated housing numbers. This is due to geographical mismatches between the different datasets used in the analysis. Most of these aggregation sets do not share exactly similar outer borders. The result is that with many of these sets, relevant gridcells are left out. This geographical mismatch problem is depicted in Figure 10. A solution for this problem is using one dataset of atomic spatial entities from which any set of zones can be amalgamated.

Figure 10 - geographical mismatch between (in this case) zones of a transportation network model ('network zones') and 4-digit postcode areas



6.4 Software used in the implemented method

The accessibility measures are calculated within Land Use Scanner, a model developed by PBL. Land Use Scanner allocates urban, industrial, natural and agricultural land use types for all of the Netherlands on a 100 meter raster. It does so from baseyear 2006 to 2040, with intermediate results in 2010 and 2020. Land Use Scanner is a comprehensive name for a set of modeling rules that are applied on a collection of data. These modeling rules are parsed, and the data is accessed and operated by means of the so-called Geo Data and Model Server (GeoDMS). According to the makers of GeoDMS, this software is a "generic software component, in use as engine for (geographic) decision and planning support systems. The GeoDMS can be configured to perform different kind of analyses on multiple data sources. The component controls all data and calculation steps" (Object Vision, 2010). Amongst other operations, GeoDMS supports network analysis methods. These have been used intensively in the presented accessibility calculations.

6.5 Improved calculation time

Initially applied calculation methods proved too costly in terms of hard disk occupation (well over 20 Gb) and calculation time (well over 2 hours on a mid-end PC). Because simulation runs (commonly runs of multiple scenarios with multiple intervals) are done repeatedly the reduction of these calculation cost is essential. The steps taken to reduce calculation costs are detailed in the next sections.

First, the amount of links in the network is reduced from approximately 1,200,000 (all roads in the Netherlands according to the Navteq dataset used in this study) to approximately 255,000 links. A common-sense approach to reduce the number of links would be to eliminate all the links classified as low-importance roads. However, the dataset's classification of roads cannot be used as basis for this reduction, because vital links in the network (e.g. ferries) were classified as low-class roads. Therefore the amount of links has been reduced by removing all links that are not used in any possible travel between the zones used in the presented method. This is done by a script in GeoDMS that first removes so-called `dangling links' iteratively; it subsequently selects all routes that are part of at least one shortest path between the origin and destination zones used in the calculations.

Furthermore, instead of creating a network for every period t, connectivity of the network is defined once, and the impedances on the network vary per period t. If at a given time a planned road is not yet open, the impedance on that stretch of road is equal to the total impedance of all other roads in the network. This saves the necessity of defining the connectivity of time-dependent networks and reduces calculation costs.

Lastly, the amount of destination zones for which accessibility is calculated has been reduced from 4,000 to approximately 1,500 zones. This is a GeoDMS-specific solution for calculation problems because of the way the shortest path algorithm is applied in GeoDMS. The GeoDMS implementation of this shortest path algorithm (a Dijkstra algorithm) calculates the total shortest path costs from all nodes in the network to the nearest origin. This implies that, when creating an exhaustive matrix of OD travel costs as in this study, the routing algorithm has to be applied repeatedly for all origins. This makes calculation costs of the shortest path calculations very dependent on amounts of origins, while calculation costs are virtually insensitive for the amounts of destinations.

Conclusively, calculation costs have been reduced by limiting the amount of network links, generating one network for all intervals and limiting the number of destination zones. With these tweaks the total calculation time of calculating accessibility for one interval, for one scenario) has been reduced to approximately 7 minutes on a mid-end PC (including the interpolation of results). Furthermore, the amount of hard disk space for one run has been reduced to approximately 600 megabytes.

7 Conclusions

This report discusses how the calculation of present and future accessibility levels can be integrated within Land Use Scanner. The resulting accessibility levels can subsequently be used as utility factors for specific land uses. Accessibility can be calculated at every interval for which Land Use Scanner allocates land use. The demonstrated exemplary results are based on travel times that are derived from interval-specific road networks. Possible effects of changes in level of congestion or road capacity are not yet taken into account.

The accessibility calculations are furthermore based on interval-specific estimations of housing and employment. These can be derived from regional projections that are also used in the definition of land use claims. In the accessibility calculations this region-based data can be disaggregated into projected amounts of housing and employment per grid cell, and subsequently aggregated into the small scale zones that are used in accessibility calculations. This disaggregation method has been tested. When aggregated to the 4-digit postcode scale that are used as zones *j* in the implemented method, the presented simple disaggregation approach yields reliable results.

7.1 Advances in this study

Several advances for Land Use Scanner applications are presented in this report. First, the inclusion of dynamic potential accessibility calculations is a step forward for Land Use Scanner. There is theoretical and empirical evidence that demonstrates the importance of social and economic opportunities in urbanization processes. Previously in Land Use Scanner, such social and economic opportunities have been proxied somewhat roughly. Furthermore, changes over time of these interaction opportunities could not be taken into account. With the inclusion of present and future potential accessibility measures as utility factors, Land Use Scanner uses a more detailed and time-specific indicator for social and economic opportunities. Furthermore, this introduces an element of path –dependency in the calculations because the accessibility calculations (and derived utility factors) are based on land use that is allocated in previous intervals.

7.2 Technical recommendations

Some technical problems were encountered in this study. First, aspects of the 2010 configuration complicated the accessibility calculations. This 2010 configuration simulates policy variants that apply land use claims at different levels of spatial aggregations. The spatial scale of projected housing and employment used for disaggregation in the accessibility calculations ought to vary in conjunction; however, this proved too complex to integrate within the time constraints of this study.

Second, when aggregating or disaggregating, geographical mismatches between different zonal datasets become apparent. These mismatches are problematic in cases where gridcell level data are aggregated to zones with non-matching outer borders. A solution for this problem is recommendable. Preferably such a solution introduces *atomic* spatial entities, from which any potential set of zones can be amalgamated without unintended geographical mismatches.

7.3 Recommendations for further research

The integrated method has some shortcomings. In this section possible extensions and further research are suggested.

First, only passenger car transport is taken into account in the implemented accessibility calculations. Next to car transport, bicycling and public transportation are important transport modes in the Netherlands, and interaction opportunities served by these modes are likely to affect land use changes as well. The presented method should therefore include these transport modes.

Next, congestion direly affects commuting times for a large share of the Dutch population. In Land Use Scanner congestion can be proxied by modelling levels of transportation demand, and relating these demands to observed travel time losses on road segments. Such an approach might therefore be tested as a basis for the inclusion of congestion effects in the accessibility calculations.

Furthermore, the presented calculations are based on only two distance-decay functions that describe activities from the home. These are particularly suitable for utility factors of residential locations. However, how accessibility affects different types of residential land uses is not yet known, and deserves work. Next, it is to be expected that actors concerned with different land uses have different valuations of travel costs. Even more, it is to be expected that the valuation of travel costs varies over time (and possibly even per scenario). Conclusively, it is recommendable to research how accessibility affects different residential land uses, what accessibility measures and distance decay functions are suitable for (amongst others) commercial and recreational land uses, and how the applied distance decay functions might change over time.

For the presented method, future projections of housing and employment are disaggregated. The disaggregation method has been tested on the current spatial distribution of housing and proves to be fairly reliable. However, systematic errors in the results of the disaggregation of current housing are apparent. The model can presumably be improved by using other factors, such as current housing densities. Furthermore, some trends in the spatial distribution of housing as well as housing densities over time might become apparent in the errors of the disaggregation method. In short, the housing disaggregation method needs further study. Even more, performance of the disaggregation method has not even been tested yet for employment.

Lastly, the presented method does not take socio-economic opportunities in bordering countries (i.e. Germany and Belgium) into account, nor does the data used in the method include foreign network links that might affect economic opportunities in border regions of the Netherlands. The influence of cross-border socio-economic opportunities and the influence of foreign network links on accessibility deserve attention in further research.

Acknowledgements

The author wishes to thank Eric Koomen (VU University Amsterdam), Bart Rijken (PBL), Auke Vlonk (PBL), Martin van der Beek (Object Vision) and Maarten Hilferink (Object Vision) for their kind advice and assistance. The author further wishes to thank PBL for supplying the necessary funding and data for this work.

References and data

Alonso, W.A. (1964) Location and land use: toward a general theory of land rent. Harvard University Press, Cambridge.

Anselin, L. (1995) Local Indicators of Spatial Association - LISA. Geographical Analysis 27 (2).

Atzema, O., Korteweg, P., Lambooy, J. and Van Oort, F. (2009) Drijvende krachten achter ruimtegebruik van werken. Universiteit Utrecht, Utrecht.

Borzacchiello, M.T., Nijkamp, P. and Koomen, E. (2010) Accessibility and urban development: a grid-based comparative statistical analysis of Dutch cities. Environment and Planning B 37(1): 148-169.

Borzacchiello, M.T., Nijkamp, P. and Scholten, H.J. (2009) A logistic regression model for explaining urban development on the basis of accessibility: A case study of Naples. International Journal of Environment and Sustainable Development 8(3/4): 300-313.

De Bok, M. and Van Oort, F. (2007) On how to analyse accessibility, agglomeration as location factors: a life cycle analysis of relocating firms in the Netherlands. 9th Nectar Conference. 2007.

De Graaff, T., Rietveld, P. and Debrezion, G. (2007) De invloed van bereikbaarheid op vastgoedwaarden. Colloquium Vervoersplanologisch Speurwerk. 2007.

Geurs, K. and Ritsema van Eck, J. (2003) Accessibility evaluation of land-use scenarios: the implications of job competition, land-use, and infrastructure developments for the Netherlands. Environment and Planning B 30(1): 69-87.

Hansen, W.G. (1959) How accessibility shapes land use. Journal of the American Institute of Planners 25: 73-76.

Hilbers, H.D. and Verroen, E.J. (1993) Het beoordelen van de bereikbaarheid van lokaties: Definiëring, maatstaven, toepassing en beleidsimplicaties. TNO, Afdeling Verkeer en Vervoer, Delft. [Accessed: 1993].

Hilferink, M. and Rietveld, P. (1999) Land Use Scanner: An integrated GIS based model for long term projections of land use in urban and rural areas. Journal of Geographical Systems 1(2): 155-177.

Horner, M.W. and Murray, A.T. (2002) Excess commuting and the modifiable areal unit problem. Urban Studies 39(1): 131-139.

Koomen, E., Loonen, W. and Hilferink, M. (2008) Climate-change adaptations in land-use planning; a scenario-based approach. In: Bernard, L., Friis-Christensen, A. and Pundt, H. (eds.), The European Information Society; Taking Geoinformation Science One Step Further. Springer, Berlin, pp. 261-282.

Loonen, W. and Koomen, E. (2009) Calibration and validation of the Land Use Scanner allocation algorithms. PBL-report. Netherlands Environmental Assessment Agency, Bilthoven.

Object Vision (2010) Object Vision Products. Available at: http://www.objectvision.nl [Accessed: 1-7-2010].

Priemus, H. and Hoekstra, J. (2009) Drijvende krachten achter woningvraag en ruimtegebruik wonen. Onderzoeksinstituut OTB, Delft.

Rietveld, P. (1989) Employment effects of changes in transportation infrastructure: methodological aspects of the gravity model. Papers of the Regional Science Association 66: 19-30.

Wegener, M. and Fürst, F. (1999) Land-Use Transport Interaction: State of the Art. Fakultät Raumplanung, Universität Dortmund, Dortmund.

Zondag, B. and Pieters, M. (2005) Influence of accessibility on residential location choice. Annual meeting of the Transportation Research Board. 2005.

Data

CBS, VROM, PBL, RIGO (2010) *Number of houses per postcode in the Netherlands, 2006*. Data originally created by Netherlands Statistics (CBS), and edited by Ministry of Housing, Spatial Planning and the Environment (VROM), RIGO and Netherlands Environmental Assessment Agency (PBL).

LISA (2006) *Number of employees per establishment, 2006*. Data edited by Chris Jacobs for this application.

Navteq (2007) *Road network of the Netherlands*. Data edited by Chris Jacobs for this application.

NIROV (2010) *Nieuwe kaart van Nederland*. Available at: http://www.kaart.nieuwekaart.nl/?page_id=13 [accessed 01-02-2010]

Statline (2010) *Woningvoorraad op 31 december 2006*. Available at: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=71477NED&D1=0&D2=0&D3=59&VW=T [accessed 08-06-2010]

SPINlab Research Memoranda

Next to this report, other background studies related to land use simulation in the LUMOS toolbox have been published as SPINlab Research Memoranda. Relevant memoranda are mentioned below. For a complete list of land-use related SPINlab publications, see www.feweb.vu.nl/gis.

Koomen, E. (2002), *De Ruimtescanner verkend; kwaliteitsaspecten van het informatiesysteem Ruimtescanner*, SPINlab Research Memorandum SL-01, Vrije Universiteit/Ruimtelijk Planbureau, Amsterdam/Den Haag.

Koomen, E., T. Kuhlman, W. Loonen & J. Ritsema van Eck (2005), *De Ruimtescanner in 'Ruimte voor landbouw'; data- en modelaanpassingen*, SPINlab Research Memorandum SL-02, Vrije Universiteit, Amsterdam.

Loonen, W., E. Koomen, P. Verburg & M. Kuijpers-Linde (2006), *Land Use MOdeling System (LUMOS): A Toolbox for Land Use Modeling.* SPINlab Research Memorandum SL-03, Vrije Universiteit, Amsterdam.

Riedijk, A. & R.J. van de Velde (ed., 2006), *Virtual Netherlands, Geo-visualizations for interactive spatial planning and decision-making: From Wow to Impact.* SPINlab Research Memorandum SL-04, Vrije Universiteit, Amsterdam.

Dekkers, J.E.C.& E. Koomen (2006), *De rol van sectorale inputmodellen in ruimtegebruiksimulatie: Onderzoek naar de modellenketen voor de LUMOS toolbox*, SPINlab Research Memorandum SL-05, Vrije Universiteit, Amsterdam.

Riedijk, A., R. van Wilgenburg, E. Koomen & J. Borsboom-van Beurden (2007), *Integrated scenarios of socio-economic and climate change; a framework for the 'Climate changes Spatial Planning' program*, SPINlab Research Memorandum SL-06, Vrije Universiteit, Amsterdam.

Bubeck, P. & E. Koomen (2008), *The use of quantitative evaluation measures in land-use change projections; an inventory of indicators available in land Use Scanner*, SPINlab Research Memorandum SL-07, Vrije Universiteit, Amsterdam.

Koomen, E. (2008), *De evaluatie van landschap in ruimtegebruiksimulatie,* SPINlab Research Memorandum SL-08, Vrije Universiteit, Amsterdam.

Jacobs, C.G.W. (2011) *De drijvende kracht van mobiliteit. Achtergrondstudie bij de vervaardiging van een nieuw LUMOS model,* SPINlab Research Memorandum SL-09, Vrije Universiteit, Amsterdam.

Other relevant LUMOS background reports

Dekkers, J.E.C. (2005), *Grondprijzen, geschiktheidskaarten en instelling van parameters in het ruimtegebruiksimulatiemodel Ruimtescanner - Technisch achtergrondrapport bij het Project Ruimtelijke Beelden*, MNP rapport, 550016005, Milieu- en Natuurplanbureau/Vrije Universiteit, Bilthoven/Amsterdam.

Loonen, W. & Koomen, E. (2009) *Calibration and validation of the Land Use Scanner allocation algorithms*, PBL publication number 550026002, Netherlands Environmental Assessment Agency (PBL), Bilthoven.