

Interaction of Transportation and Land Use: A Simple Model for Land Use Allocation and Transportation Demand

Rolf Moeckel
Lead Research Engineer
PB
One Penn Plaza
New York, NY 10119
Tel. +1.212.465.5630
moeckel@pbworld.com

Patrick J. Costinett
Senior Professional Associate
PB
999 Third Avenue, Suite 2200
Seattle, WA 98104
Tel. +1.206.382.5200
costinett@pbworld.com

Tara J. Weidner
Project Manager
PB
400 SW Sixth, Suite 802
Portland, OR 97204
Tel. +1.503.274.8772
weidner@pbworld.com

Submitted on August 1, 2007; revised on November 15, 2007, 5,886 words, 5 figures

Abstract

Most transportation modelers agree that the integration of land use is beneficial for transportation simulation. As travel demand derives from the distribution of land uses and as traffic flows shape land use by accessibilities, integrated land-use/transportation models improve transportation demand forecasts. However, there are only very few transportation models that actually integrate with land use simulations. Commonly, land uses remain unchanged over time or exogenous forecasts are used to determine population and employment distributions in future years. This paper presents an easy-to-use alternative of a land use model simulating population, employment, demand for commute trips and demand for goods flows. Using a series of Iterative Proportional Fitting (also called Matrix Balancing) it runs quickly and follows straightforward calibration procedures. The model offers "pure" transportation models a simple alternative to become an integrated transportation/land-use model.

Acknowledgement

The design and implementation of this model was funded by PB under the direction of Rick Donnelly and received strong support from Kimberly Grommes and Christi Willison from the PB office in Albuquerque, NM. Based on his comprehensive expertise in urban modeling, Rick gave valuable incites and feedback for the model design. Kimberly and Christi were strongly engaged in transforming the model design into a powerful application tool for urban simulation. The authors are grateful for using a large amount of input data from the Ohio Statewide Modeling Project (OSMP).

1 Introduction

Travel behavior and land use changes are linked closely. Depending on where people live, work, shop and spend their leisure time travel demand arises. Likewise, the location of firms shapes the commodity flows. On the other hand, the transportation system and in particular travel times from one location to other destinations also have a strong impact on land use. Those areas that are more easily accessible become more attractive for both population and firms. The integration of a transportation simulation with land-use modeling allows simulating the entire land-use/transportation feedback cycle.

Simulation of land use in mathematical models has a history of half a century. One of the first approaches was the model of Herbert and Stevens (1960) in cooperation with Britton Harris (1966) simulating the distribution of households to residential land use as an equilibrium model. Another pioneering example is the Model of Metropolis by Lowry (1964) allocating firms of the basic sector with the housing location of their employees and solving firms of the local sector and their employees iteratively. Based on this work of Lowry and on the famous work on urban dynamics by Forrester (1969) many land use models have been developed in the following decades. The most prominent examples are MEPLAN (Echenique et al. 1969, Echenique et al. 1990), PLUM (Rosenthal et al. 1972, Goldner et al. 1972, Reynolds and Meredith 1972) which was later replaced by DRAM/EMPAL (Putman 1983), TRANUS (de la Barra 1989: 143 ff, de la Barra et al. 1984), IRPUD (Wegener 1998a, 1982), MUSSA (Martínez 1996), DELTA (Simmonds 2001, 1999), California Urban Futures Model (Landis and Zhang 1998a, 1998b), ALBATROSS (Arntze and Timmermans 2000), PECAS (Hunt and Abraham 2003), UrbanSim (Waddell et al. 2003, Waddell 2002), and ILUMASS (Moeckel et al. 2007, Strauch et al. 2005).

Extensive overview on current land-use/transportation models can be found in Hunt et al. (2005), Wegener (2004, 1998b, 1994), Timmermans (2003), Kanaroglou and Scott (2002), the U.S. Environmental Protection Agency EPA (2000: 27 ff), and Kain (1987). These summaries show that, recently, the tendency has been to develop ever more complex simulation models. Many model developers have moved towards microsimulation models that simulate individuals and their interaction rather than aggregate groups. While the number of ideas simulating detail and complexity has grown quickly, only few models have actually reached the status of a working, calibrated and validated simulation tool. The majority of recent land-use models remain sophisticated academic research tools rather than applicable planning support systems. This is one of the main reasons why current transportation models tend to be developed without land use simulations.

This lack of reliable ready-to-apply land use models motivated the development of SEAM (Simple Economic Allocation Model). The major driver was to develop a model that is easy to use, readily populated with data and reliable in terms of simulation results. Thus, practicability was valued higher than academic search for detail.

The land-use simulation SEAM has been developed to provide an uncomplicated land-use model that can be integrated with any transportation simulation (Figure 1). The Transportation Simulation provides travel times, distances and/or costs, which are transformed into accessibilities by SEAM. Depending on these updated accessibilities and regional control totals for population and employment, SEAM updates the location of households and jobs. Additionally, SEAM derives flows of labor (i.e. commuter trips) and flows of commodities (i.e. goods flows). Flows are provided as an origin/destination matrix that can be assigned to the transportation network by the transportation model. The spatial resolution is based on zones.

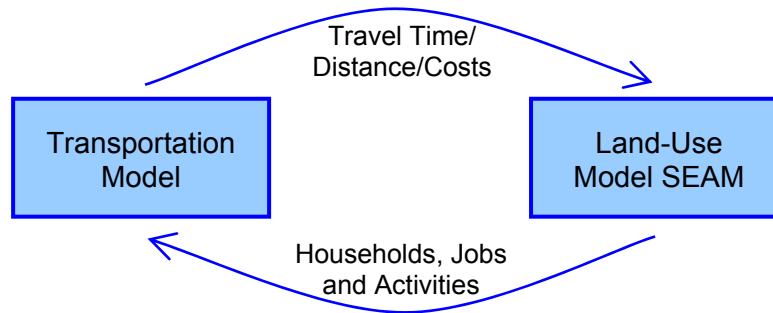


Figure 1: Interaction of a transportation model with SEAM

A major goal for the development of SEAM is to develop a model that is easy to implement, to calibrate and to use. Input data are limited to the required minimum, and sophisticated detail is replaced by appealing simplicity. The calibration is limited to straightforward statistical procedures such as linear regression and iterative proportional fitting. The model focuses on providing the required data used by the transportation simulation.

2 Data Requirements

SEAM was initially developed for the study area of the Ohio Statewide Modeling Project (OSMP) and draws much of the necessary data from that project. The area consists of the entire state of Ohio and adjacent counties in the states of Pennsylvania, West Virginia, Kentucky, Indiana and Michigan. The entire area has a population of approximately 13.9 million living in 5.5 million households and 8.3 million jobs. The study area is subdivided into 4248 Transportation Analysis Zones (TAZ) forming the zones for the simulation.

Data drawn directly from the OSMP project include roadway and transit networks, socioeconomic data by TAZ, land use by TAZ and forecasts of future population and employment by region and county. Additional data developed specifically for SEAM utilizing Census PUMs data include: households by household type by TAZ, average household size by household type, average workers per household by household type, workers by occupation by residence TAZ and workplace TAZ, average income per worker by household type, and occupation versus industry cross-classifications. County-to-county labor flows were also developed from Census CTPP Journey-to-Work data for the model area and between the model area and external residences and workplaces.

Additionally, SEAM uses data that is simulated within the OSMP model every simulation period. A crucial input is the residential and non-residential floorspace inventory. Furthermore, travel time, travel costs and mode choice logsums are used as provided by the OSMP transportation model.

3 Model Components

The SEAM approach is built as a doubly-constrained gravity model to allocate employment and households based on utilities of zones for different users. The model consists of five major parts. Initially, accessibilities are calculated to integrate the feedback from transportation on land use. Population, or more precisely households, are then allocated depending on these accessibilities and other factors. Based on the updated location of population and the accessibilities, employment in the form of jobs is allocated. The fourth module connects population with employment by estimating commute trips of workers from their home to their work location. Finally, flows of commodities are simulated based on the location of employment and the commodity production and consumption rates. In its current application, SEAM is integrated into the OSMP project. SEAM uses floorspace provided by a developer model and mode choice logsums as well as travel times provided by the transportation model within OSMP.

SEAM distinguishes twelve household types, defined by two household size classes and eight total household income classes. Table 1 lists the household type definitions.

Name	Household Size	Annual Household Income in US\$
HHlt20K1to2	1 – 2	less than 20,000
HHlt20K3plus	3 or more	
HH20to40K1to2	1 – 2	20,000 – 40,000
HH20to40K3plus	3 or more	
HH40to60K1to2	1 – 2	40,000 – 60,000
HH40to60K3plus	3 or more	
HH60to75K1to2	1 – 2	60,000 – 75,000
HH60to75K3plus	3 or more	
HH75to100K1to2	1 – 2	75,000 – 100,000
HH75to100K3plus	3 or more	
HH100Kplus1to2	1 – 2	more than 100,000
HH100Kplus3plus	3 or more	

Table 1: Definitions of household types

On the employment side 23 industries are distinguished, namely Agriculture Forestry and Fisheries, Primary Metal Products, Light Industry, Heavy Industry, Transportation Equipment, Wholesale, Retail, Hotel and Accommodation, Construction, Health Care, Transportation Handling, Utilities Services, Other Services, K-12 Education, Post-Secondary Education, and Government and Other. The first seven industries are subdivided into office-related employment and production employment to better reflect locational tendencies.

3.1 Accessibility

Accessibility is an aggregate measure for each zone describing how many activities can be reached how easily. Accessibility increases with the number of available activities and decreases with the distance of these activities from a given zone. The importance of accessibility for urban development was described early by Hansen who proved statistically the correlation between im-

provements in accessibility and new urban development in Washington D.C. (Hansen 1959). Kreibich demonstrated the impact of light rail construction and improved accessibility for the Olympic Games in Munich on residential development (Kreibich 1978). Different approaches to calculate accessibility indicators have been developed; an overview is given by Schürmann et al. (1997).

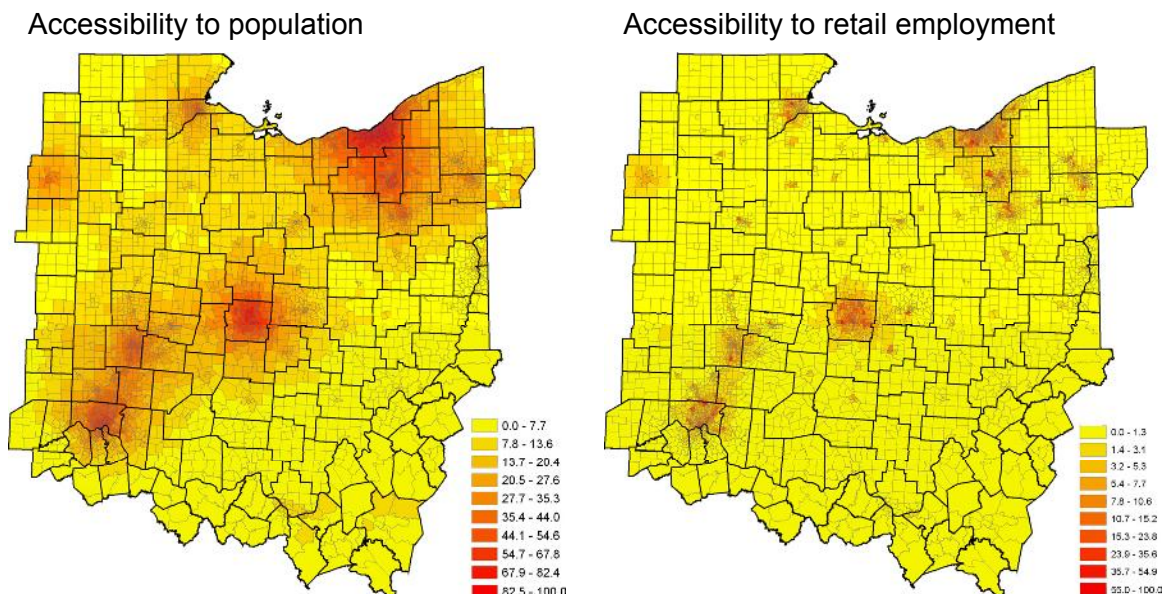
In SEAM, the potential accessibility (also called Hansen accessibility) is applied because it relates both to relevant activities to be reached and associated travel impedance, which is measured by travel time, distance, costs, or a combination thereof. SEAM applies impedance in the form of mode choice logsums by the transportation model. These logsums combine the travel times by various modes into a single measurement. The weighting of travel time and travel distance is distinguished by three income groups. Equation 1 shows how the potential accessibility used in SEAM is calculated.

$$a_{i,k} = \sum_j w_{j,k}^{\alpha} \cdot \exp(m_{ij}^{\beta}) \quad (1)$$

- with $a_{i,k}$ Accessibility from zone i for activity k
 $w_{j,k}$ Number of activities k in zone j
 m_{ij} Mode choice logsum between zone i and zone j
 α Coefficient to increase weight of larger agglomerations (cluster effect)
 β Coefficient to set weight of impedance factor m_{ij}

SEAM estimates accessibilities of different destinations, representing the myriad of activities population and firms seek for in their location decisions. Accessibilities of 25 destinations are calculated, among those the accessibility of total population, family households and jobs in selected industries.

The parameters α and β for equation 1 are set heuristically. Based on modeling experience and plots of accessibilities for the study area the agglomeration parameter α is set to 1.5 for every accessibility measure (values above 1 increase the importance of larger agglomerations). The β parameter has been set after several trials to 0.5 for accessibilities of total population and of family households and 1 for accessibilities of every employment type. This larger employment factor is based on the assumption that firms have a tendency to cluster closely to other firms of the same



type. Setting a higher coefficient for β reduces the importance of employment that is further away and stresses the importance of nearby employment.

The accessibility values bear no unit. To keep accessibilities to different destinations comparable they are normalized to values between 0 and 100. As an example, figure 2 shows maps of the study area colored according to the calculated zonal accessibility to total population and to retail employment. It shows that retail facilities are more clustered than population, reflected in the more dispersed accessibility pattern of population than of retail employees. This effect is emphasized by appropriate values for α and β .

Figure 2: Accessibilities to population (left) and to retail employment (right)

3.2 Population Distribution

The second main component is the allocation of population. As SEAM distinguished household types, the result of this module is a matrix with number of households by 16 types and by 4248 zones. The attractiveness of every zone for each household type is estimated by a utility function. Regional control totals are households by type and total households by county. Households by type are given by the OSMP Synthetic Population Generator, and households by county are based on county population forecasts by Ohio Department of Development (Ohio DOT) and other agencies in adjacent states. Matrix balancing, first described by Deming and Stephan (1940) and also known as Iterative Proportional Fitting or RAS method, is used to distribute households in a matrix in such a way that the number of households in each matrix cell are close to observed values and satisfy the regional control totals.

The seed matrix for the matrix balancing is filled with utilities for households by type to locate in that particular zone. The utilities are estimated for each household of a given type and every zone based on total available residential floorspace from the OSMP Floorspace Model, total employment, employment in retail facilities, accessibility to total population, accessibility to post secondary education employment and accessibility to health care employment. Equation 2 is used calculate the utility.

$$u_{i,h} = c + \alpha_1 \cdot F_{Res,i} + \alpha_2 \cdot E_{Tot,i} + \alpha_3 \cdot E_{Ret,i} + \alpha_4 \cdot A_{Pop,i} + \alpha_5 \cdot A_{Edu,i} + \alpha_6 \cdot A_{Hea,i} \quad (2)$$

with $u_{i,h}$	Utility of zone i for household type h
$F_{Res,i}$	Residential floorspace in zone i (current simulation period)
$E_{Tot,i}$	Total Employment in zone i (previous simulation period)
$E_{Res,i}$	Employment in retail facilities in zone i (previous simulation period)
$A_{Pop,i}$	Accessibility of population for zone i
$A_{Edu,i}$	Accessibility of employment in post secondary education for zone i
$A_{Hea,i}$	Accessibility of employment in health care for zone i
c, α	Coefficients, to be estimated for every household type h

The calibration of the constant c and the coefficients α_1 to α_6 is described in section 4. The utilities describe the likelihood that a household of a particular type would choose to locate in a particular zone.

The population forecast provided by Ohio DOT is given at the county level and has to be disaggregated to TAZ zones used in SEAM. Equation 3 is used to estimate the number of households by TAZ.

$$HH_i = HH_{County(i)} \cdot \frac{\sum_h u_{i,h}}{\sum_{j \in County(i)} \sum_h u_{j,h}} \quad (3)$$

with HH_i Number of households in zone i
 $HH_{County(i)}$ Number households in the county of zone i
 $u_{j,h}$ Utility of zone j for household type h

The sum of utilities in every TAZ is used to disaggregate the county population forecast to zones, i.e. zones with a higher total utility receive a larger share of households than zones with a lower utility.

Table 2 visualizes the settings for the population matrix balancing operation. The utilities $u_{i,h}$ for each household type for every zone are calculated, the sum of model wide households by type is given exogenously, households by county are given by the exogenous population forecast and disaggregated to TAZ by the sums of utilities within every zone.

TAZ	Household type 1	Household type 2	...	Household type 12	Households by TAZ	Households by County
1	$u_{i,h} = 0.253$	$u_{i,h} = 0.321$		$u_{i,h} = 0.051$	5,295	
2	$u_{i,h} = 0.024$	$u_{i,h} = 0.814$		$u_{i,h} = 0.135$	6,018	16,823
3	$u_{i,h} = 0.325$	$u_{i,h} = 0.435$		$u_{i,h} = 0.653$	5,510	
...						
5002	$u_{i,h} = 0.461$	$u_{i,h} = 0.467$		$u_{i,h} = 0.671$	1,494	
Households by type	1,013,855	251,857		297,774	5,536,199	5,536,199
Seed Matrix						

Table 2: Seed matrix and column/row sums to estimate households by zone and type

The matrix balancing adjusts the cells of the matrix iteratively by rows and by columns attempting to match the row and column totals. The adjusting stops when a convergence criteria is met. This typically is set as a minimum threshold error between the model's estimated row and column totals and the exogenous regional control totals, or if a given maximal number of iterations has been reached. In an average model run a relative error below 0.001 is reached within two iterations.

3.3 Employment Distribution

The employment distribution module is similar to the population module. Matrix balancing is used to adjust a matrix filled with utilities for different industries for every zone. First, the utilities for every zone distinguished by industry are estimated.

$$u_{i,f} = c + \alpha_1 \cdot L_{1,i} + \alpha_2 \cdot L_{2,i} + \alpha_3 \cdot L_{3,i} + \dots + \alpha_n \cdot L_{n,i} \quad (4)$$

with $u_{i,f}$ Utility of zone i for firms in industry f
 $L_{n,i}$ Location factor n in zone i , including:
 - Accessibility to employment in various industries
 - Accessibility to population
 - Availability of floorspace of various types
 - Population density
 - Total employment
 c, α Coefficients, to be estimated for every industry f

Not all of the 33 location factors f_1 to f_{33} is relevant for each industry. A careful selection of location factors that are theoretically meaningful and statistically significant is described in section 4 along with the description of the calibration process for the constant c and the coefficients α_1 to α_n . The utilities $u_{i,f}$ describe the likelihood that a firm of a particular industry would choose to locate in a particular zone.

The employment forecast from Ohio DOT is used as a region control total for the entire study area and has to be disaggregated to TAZ that are used in SEAM. Equation 5 is used to estimate the number of jobs by TAZ.

$$E_i = E_{Forecast} \cdot \frac{\sum_f u_{i,f}}{\sum_j \sum_f u_{j,f}} \quad (5)$$

with E_i Number of jobs in zone i
 $E_{Forecast}$ Exogenous forecast of total number of jobs
 $u_{i,f}$ Utility of zone i for business type f

The sum of utilities for all business types is used to disaggregate the total employment forecast to zones, i.e. zones with a higher utility over all business types receive a larger share of jobs than zones with a lower utility.

Regional employment control totals for the 23 industry categories described above are provided by the OSMF Interregional Economic Model. Given the number of jobs per TAZ and number of jobs per industry type, matrix balancing is used to estimate the number of jobs per zone and industry type. Table 3 visualizes the procedure.

TAZ	Industry 1	Industry 2	...	Industry 23	Jobs by TAZ
1	$u_{i,f} = 0.465$	$u_{i,f} = 0.046$		$u_{i,f} = 0.066$	1,681
2	$u_{i,f} = 0.919$	$u_{i,f} = 0.794$		$u_{i,f} = 0.956$	4,917
3	$u_{i,f} = 0.154$	$u_{i,f} = 0.494$		$u_{i,f} = 0.297$	1,985
...					
5002	$u_{i,f} = 0.479$	$u_{i,f} = 0.307$		$u_{i,f} = 0.823$	615
Jobs by type	108,776	77,413		328,227	8,302,143
					Seed Matrix

Table 3: Seed matrix and column/row sums to estimate jobs by zone and type

Rows and columns of the seed matrix are adjusted iteratively to meet the marginal totals. The process stops when the convergence criteria is met, again either a minimum difference with mar-

ginals or a maximum number of iterations. In an average model run a relative error below 0.001 is reached after six iterations.

3.4 Labor Flows

The SEAM Labor Flow module estimates intra- and interzonal commute trips distinguished by occupation type. Based on the results of the Population (section 3.2) and the Employment (section 3.3) modules, home and employment locations of workers are revealed. A gamma function using mode choice logsums is applied to set the impedance for work trips. 18 occupation types are distinguished:

The utility, or more precisely the negative utility or impedance to travel from zone i to zone j , is calculated by a gamma function using mode choice logsums from the transportation model, shown in equation 6. The gamma function has been chosen as this curve tends to represent well distributions of commute distances.

$$u_{o,ij} = m_{o,ij}^{\beta} \cdot \exp(-\gamma \cdot m_{o,ij}) \quad (6)$$

with $u_{o,ij}$ Utility for a worker with occupation o to commute from zone i to j
 $m_{o,ij}$ Mode choice logsum for occupation o from zone i to zone j
 β, γ Parameters, to be estimated for every occupation type

The origins and destinations of the commuter flow matrix are the housing locations of workers (origin) and the locations of jobs (destinations). The results of the SEAM module for population (section 3.2) are used to estimate housing location of workers. As SEAM estimates number of households by type in every TAZ these numbers have to be translated into number of workers by zone. Equation 7 is applied to estimate workers based on households. The number of households of each type in each zone is multiplied by the average number of workers in this household type and the average share of occupation o in this household type, obtained from Census PUMS data.

$$W(\text{home})_{i,o} = \sum_h HH_{h,i} \cdot e_h \cdot os_{o,h} \quad (7)$$

with $W(\text{home})_{i,o}$ Number of workers at home location for occupation o in zone i
 $HH_{h,i}$ Number of households of type h in zone i
 e_h Average number of workers in households of type h
 $os_{o,h}$ Occupation share of occupation o for household type h

The number of jobs by occupation at each work location zone is estimated based on the number of jobs simulated by the employment module of SEAM (section 3.3). Equation 8 shows how jobs as a destination of work trips are estimated. The number of jobs is multiplied with a factor to translate jobs into workers and the average share of occupation o in this industry. The workers-per-job factor addresses a common discrepancy between employment data and demographic data and provides jobs as full-time equivalent workers. Quite a few jobs are filled by two (or more) part-time workers, thus jobs have to be translated into workers to match the right number of work trips.

$$W(\text{work})_{i,o} = \sum_f E_{f,i} \cdot wj_f \cdot os_{o,f} \quad (8)$$

with $W(\text{work})_{i,o}$ Number of workers at work location for occupation o in zone i

E_{fi}	Number of jobs in industry f in zone i
wj_f	Factor of workers per job in industry f
os_{of}	Occupation share of occupation o for industry f

After calculating the origins and destinations of work trips for each occupation type the matrix balancing can be applied to estimate commute trips (table 4). Again, the convergence criterion is a minimum error relative to the marginals, or a maximum number of iterations.

TAZ	1	2	3	...	5002	Jobs at work (destination)
1	$u_{o,ij} = 0.95$	$u_{o,ij} = 0.31$	$u_{o,ij} = 0.01$		$u_{o,ij} = 0.06$	160
2	$u_{o,ij} = 0.49$	$u_{o,ij} = 0.98$	$u_{o,ij} = 0.68$		$u_{o,ij} = 0.63$	24
3	$u_{o,ij} = 0.48$	$u_{o,ij} = 0.61$	$u_{o,ij} = 0.91$		$u_{o,ij} = 0.28$	117
...						
5002	$u_{o,ij} = 0.04$	$u_{o,ij} = 0.15$	$u_{o,ij} = 0.02$		$u_{o,ij} = 0.96$	61
Workers at home (origin)	32	8	512		91	6,529
Seed Matrix						

Table 4: Seed matrix and column/row sums to estimate commute flows by occupation

The labor flow matrix can be used by the transport model to simulate work traffic flows. Different occupation types allow simulating mode choice in accordance to the likelihood of certain occupation types to use specific modes.

In this context, labor flows are the only personal travel flows that are simulated. It would be possible to simulate other trip purposes, such as educational trips or retail trips. However, an important limitation of this approach is that only one-purpose trips can be estimated, but no tour-based trips that add purposes such as a home-work-shopping-home trip. A distinction about modes chosen has to be done by a transportation model, the purpose of this Labor flow module in SEAM is to provide estimates of the likely distribution of work trips.

3.5 Commodity Flows

The SEAM module for commodity flows estimates flows of goods movement. Flows are measured in dollars. Again, matrix balancing is used to adjust a seed matrix until the rows and columns match marginals. The distance d_{ij} between zones for truck trips is used as the seed matrix and transformed by a gamma function (equation 9).

$$u_{c,ij} = d_{ij}^{\beta} \cdot \exp(-\gamma \cdot (d_{ij})^s) \quad (9)$$

with $u_{c,ij}$	Utility to transport commodity c from zone i to zone j
d_{ij}	Distance for truck trips from zone i to zone j
s	Scaling factor to increase weight of long distance commodity flows
β, γ	Parameters, to be estimated for every commodity type

In contrast to labor flows, the equation for commodity flows includes a scaling factor that is used to increase the weight of the very distant flows. The shape of the gamma function tends to provide probability values close to 0 for large distances. This works well for labor flows, as there is a

natural limit of how many miles a commuter may travel on a daily basis. Commodity flows, however, may move over long distances. The scaling factor applied in equation 9 keeps the overall shape of the gamma function but delays the decline to 0, thus long-distance flows may be simulated.

The origins and destinations of good's flows are derived from the locations of jobs. Equations 10 and 11 are applied to calculate the target values. The second term of in both equations represents the quantity of various goods commodities made and used by each industry, using existing technologies.

$$CO_{i,c} = \sum_f E_{f,i} \cdot p(in)_{c,f} \frac{Act_f}{\sum_i E_{f,i}} \quad (10)$$

$$PR_{i,c} = \sum_f E_{f,i} \cdot p(out)_{c,f} \cdot \frac{Act_f}{\sum_i E_{f,i}} \quad (11)$$

with $CO_{i,c}$ Consumption (target) of commodity c in zone i in dollars
 $PR_{i,c}$ Production (origin) of commodity c in zone i in dollars
 $E_{f,i}$ Employment in industry f in zone i (provided by SEAM)
 $p(in)_{c,f}$ Consumption of commodity c by industry f for each produced dollar unit
 $p(out)_{c,f}$ Production of commodity c by industry f for each produced dollar unit
 Act_f Total activity in the study area in industry f in dollars

The origins and destinations calculated in equations 10 and 11 are used to estimate good's flows between TAZ. Table 5 shows the matrix that needs to be balanced.

TAZ	1	2	3	...	5002	Consumption (destination)
1	$u_{c,ij} = 0.90$	$u_{c,ij} = 0.37$	$u_{c,ij} = 0.03$		$u_{c,ij} = 0.26$	9,644,161
2	$u_{c,ij} = 0.19$	$u_{c,ij} = 0.89$	$u_{c,ij} = 0.63$		$u_{c,ij} = 0.70$	674,610
3	$u_{c,ij} = 0.52$	$u_{c,ij} = 0.17$	$u_{c,ij} = 0.97$		$u_{c,ij} = 0.16$	51,319,076
...						
5002	$u_{c,ij} = 0.46$	$u_{c,ij} = 0.65$	$u_{c,ij} = 0.04$		$u_{c,ij} = 0.98$	1,347
Production (origin)	1,617,231	394,067	22,461,317		14,008	3,567,617,130
						Seed Matrix

Table 5: Seed matrix and column/row sums to estimate good's flows by commodity type

Again the matrix balancing continues until convergence criteria are met, including minimum error or maximum number of iterations.

4 Calibration

In section 3 several equations with parameters are described. With the exception of parameters for accessibility, which are estimated heuristically, the model parameters are estimated economet-

rically. The parameters are calibrated to allow the model to represent real world development. Section 5 will test whether the estimated parameters empower the model to work reliably.

4.1 Population

The calibration of the population module is done by linear multiple regression. Number of households is used as the dependent variable, and residential floorspace, total employment, retail employment, accessibility to population, accessibility to post secondary education employment, and accessibility to health care are applied as independent variables. The selection of independent variables was based on theoretical considerations and statistical significance.

The estimated parameters are differentiated by household type. A stepwise regression has been applied to include only those independent variables that bear significance for a particular household type.

As an example, the parameter estimation for household type HHlt20k1to2 (households with an income below \$ 20,000 and with a household size of 1 or 2 persons) is shown. Five out of six independent variables were found to be statistically significant at the 5 percent level (table 6). The column "Coef." shows the estimated parameter to be used in the model (c and α in equation 2). The standard error and the t-statistics show how well the variable is predicting the number of households. The column "Sign." contains the significance, in this case every variable (not the constant, which is not of further relevance here) is significant at least at the 5 percent level. The final column "Beta" explains the influence of the independent variable on the number of households of this type. While the availability of floorspace has the highest importance, retail employment has a much smaller, but still statistically significant influence.

Independent Variable	Coef.	Std. Err.	t	Sign.	Beta
Residential floorspace	0.0001	0.0000	69.28	0.0000	0.7270
Total employment	0.0110	0.0012	8.78	0.0000	0.1243
Accessibility to health care	6.1431	0.8676	7.08	0.0000	0.0728
Accessibility to post secondary education	2.7348	1.2357	2.21	0.0270	0.0236
Retail employment	0.0113	0.0056	2.00	0.0450	0.0270
Constant	1.5912	3.5574	0.45	0.6550	.

Table 6: Parameter estimation for household type "HHlt20k1to2"

The estimation of parameters has been done similarly for all twelve household types. Every type receives at least three significant independent variables as location factors. They are used to estimate the distribution of the entire population.

4.2 Employment

The allocation of jobs has been calibrated by linear multiple regression as well. Employment is distinguished by 23 split industries. A theoretical preselection of relevant location factors has been made to identify those aspects that matter to a particular industry. A list of factors that were applied include 21 accessibilities of various industries. A fairly high value for β has been chosen to calculate these accessibilities, placing a higher importance on nearby employment thus, this accessibility represents industry agglomeration effects. Since some firms seek for closeness to

population, accessibility to population is included as well. There are an additional twelve factors other than accessibilities, relating to available floorspace by type and demographics, such as population density, total number of households or employment. Certainly, a single industry does not consider all 33 location factors at the same time. Instead, a theory-based selection of relevant location factors has been done and statistically relevant ones are included in the model.

To give an example table 7 summarizes the calibration results for the industry type Transportation Equipment Production. Out of those location factors that have been selected to be theoretically consistent for this industry type, four location factors turn out to be statistically significant. The strongest influence was found to be the availability of heavy industrial floorspace, followed by accessibility of employment of the same industry type. This resembles that this industry types requires floorspace for heavy industry and tends to cluster quite strongly. The two location factors residential floorspace and population density bear a negative sign, indicating that this industry locates away from these areas, such as in industrial parks or on greenfield development.

Independent Variable	Coef.	Std. Err.	T	Sign.	Beta
Heavy industrial floorspace	0.2358305	0.004659	50.62	0	0.542612
Accessibility to TranspEquip_Prod.	24.64656	0.6346653	38.83	0	0.407125
Residential floorspace	-3.63E-06	9.97E-07	-3.64	0	-0.037849
Population density	-1.634799	0.4614518	-3.54	0	-0.036066
Constant	-19.44994	2.880927	-6.75	0	.

Table 7: Parameter estimation for industry type "Transportation Equipment Production"

Appropriate parameters have been derived similarly for all 23 industry types. Only those that are both theoretically sound and statistically significant were included in the model.

5 Validation

The calibrated model needs to be validated to test how closely the model represents real world development. Therefore, the validation process compares simulated data with real world observations as target data. Following, scatter plots are used to compare target data with model results for the year 2000. Ideally, a validation process should analyze a certain time period, for instance ten years of simulation. In this case, however, as currently developed, SEAM is highly dependent on data produced by other modules of the OSMP model. Thus, the particular influence of SEAM on the results could not be extrapolated if a decade was simulated, but rather the performance of the entire OSMP model would be observed. To analyze the capability of SEAM to represent real-world development, simulation results of the year 2000 are investigated. For this validation, input data is compiled from exogenous sources rather than from other OSMP modules. In this way, the validity of the SEAM model can be tested.

Figure 3 shows a scatter plot of the simulation of households. Every dot compares model results with target data for one TAZ in the year 2000. The overall correlation of $R^2 = 0.92$ indicates a close fit between model results and observed target data. Even though some dots deviate from the ideal diagonal line, the overall match is considered sufficient for use in transportation models. The same analysis has been done for each household type. The weakest correlation of $R^2 = 0.4803$ arises for household type HHlt20k3plus (household income below \$ 20,000 and three or more household members). This indicates the heterogeneity of this household type, with low-income families and shared housing. The highest correlation of $R^2 = 0.8967$ is achieved by

household type HH40to601to2 (household income 40,000 to 60,000 with one or two household members). Most household types reach a correlation between 0.75 and 0.85, showing a sufficient fit between model results and target data.

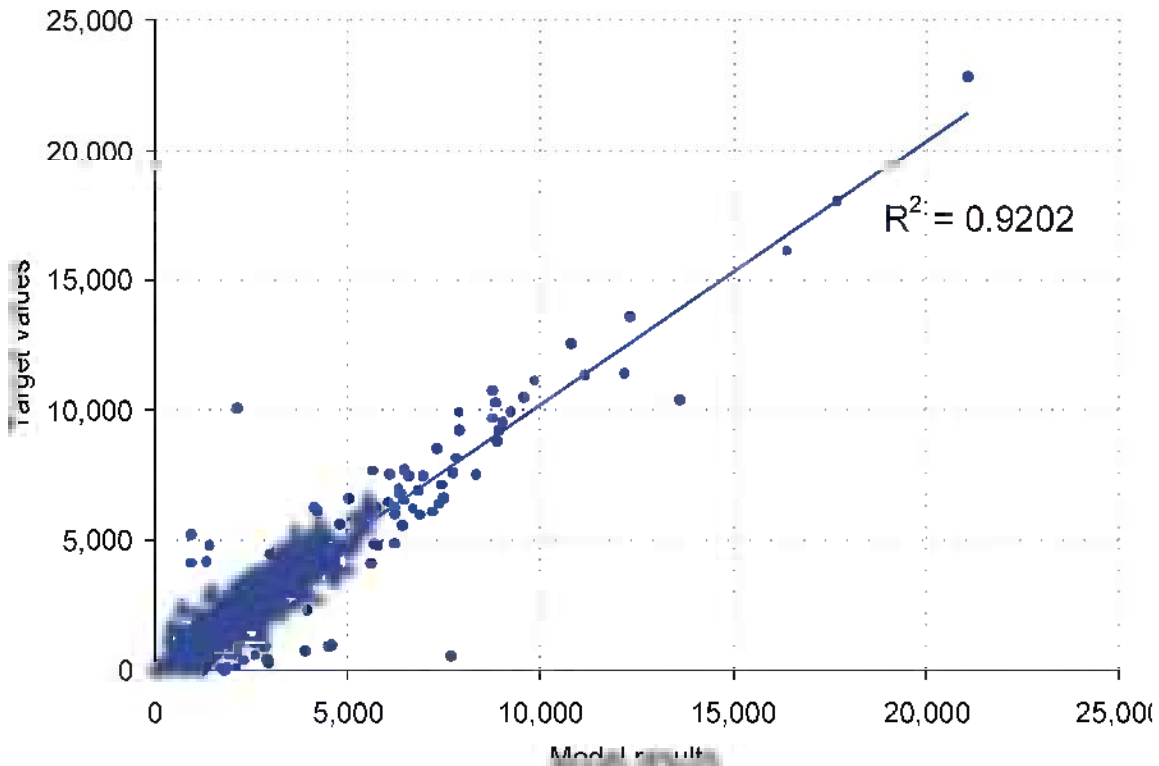


Figure 3: Model fit for total population

The same analysis has been made for the employment module. Figure 4 shows the scatter plot for total employment in the year 2000. Again, every dot represents one of the 4282 TAZ. The resulting correlation of $R^2 = 0.9757$ is even higher than for population. The poorest fit is given by jobs in Transportation Handling with a correlation of only $R^2 = 0.1239$, indicating that firms in this industry, mostly public authorities, do not follow common economic location factors. A perfect correlation of $R^2 = 1.000$ (rounded value) is achieved by Grade School Education and Agriculture, Forestry and Fishery Production, as they use floorspace that is reserved exclusively for their industry (fields/forests/fisheries or Grade School buildings), thus given the floorspace these employment types can be allocated perfectly. Most industries reach a correlation between 0.5 and 0.6, which is assumed to be sufficient as input data for transportation models.

To validate labor flows a distribution of commute distances is analyzed. There are no statistics providing labor flows by TAZ. Hence, the share of different trip distances is compared. Figure 5 shows the commute trip lengths reported in the statewide household survey in comparison with model results for occupation Managers as an example. In general, the distribution of labor flows resembles the gamma function applied in equation 6. Compared to the target values, very short trips are slightly underestimated and trips with 5 to 25 miles are slightly overestimated, but overall the fit is assumed to be sufficient. The root mean square error for commute trips of this occupation is 0.017, indicating that the distribution of trip length is well-represented. The other 17 occupation types show comparable fits between model results and target data.

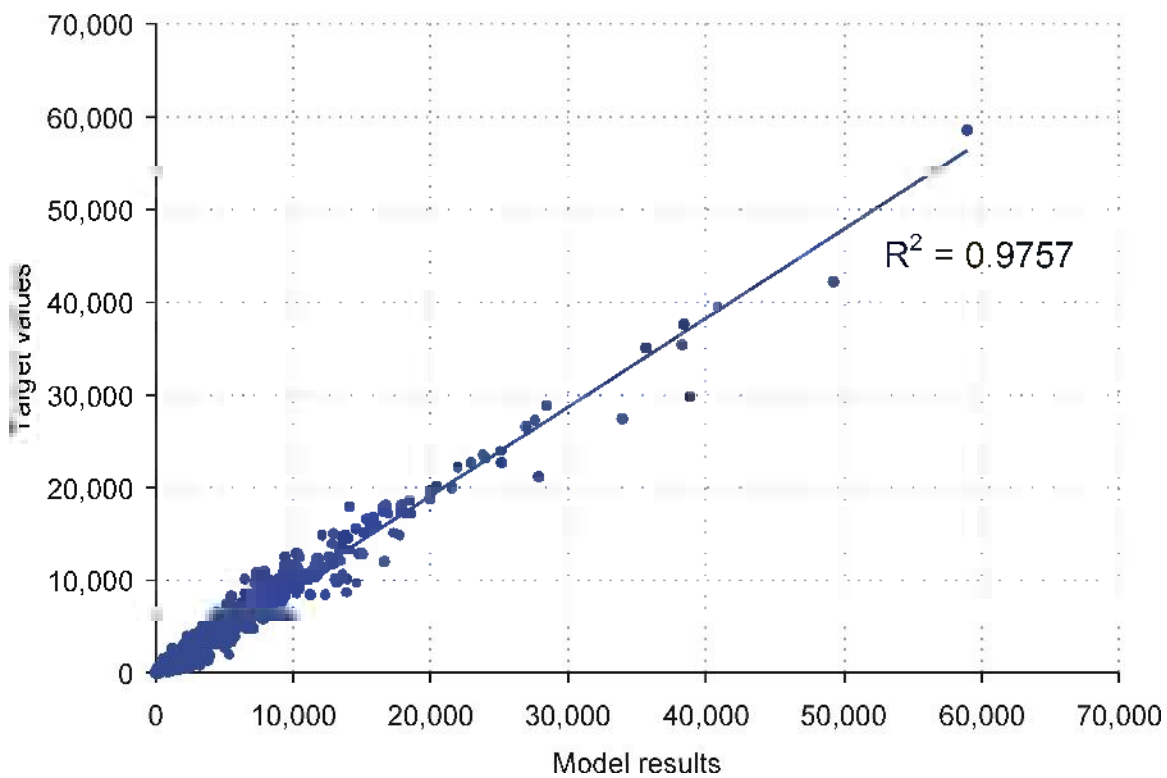


Figure 4: Model fit for total employment

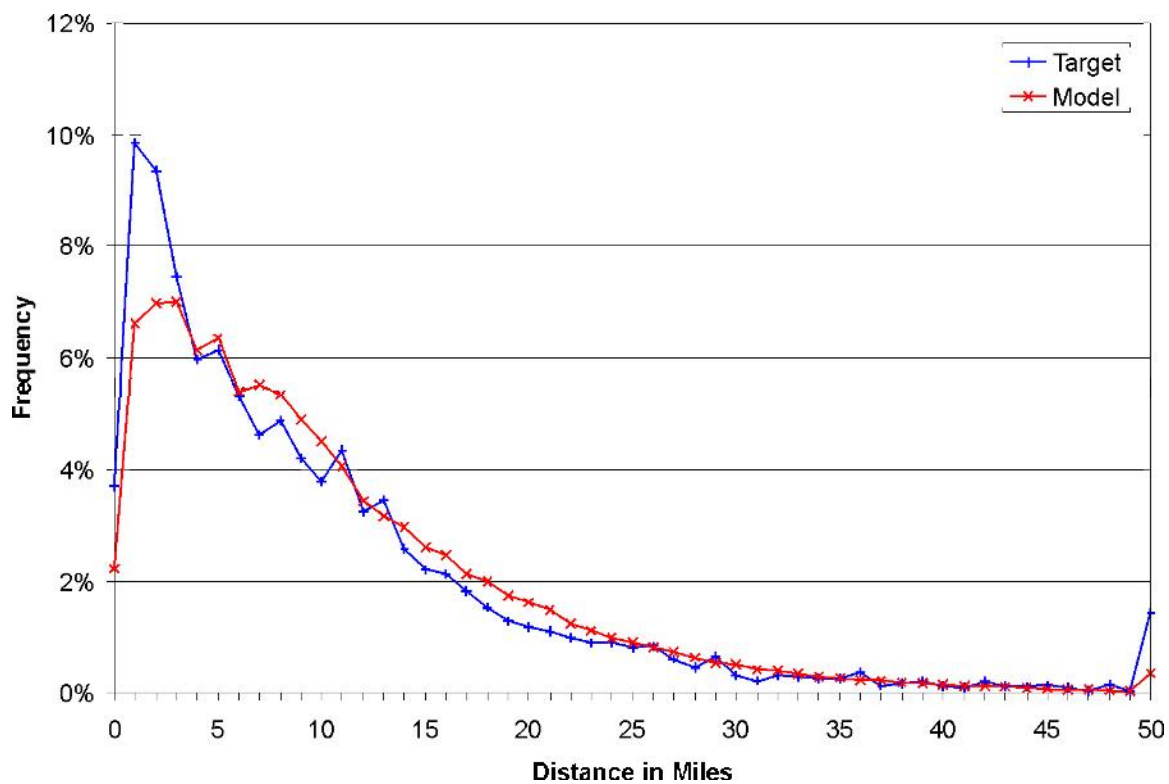


Figure 5: Model fit for labor flows of Assembly and Fabrication Workers

6 Conclusion

Because travel is derived from the location of people and businesses, a land use model is important for accurate forecasts of travel demand. The SEAM model discussed here is a simple but potentially powerful alternative. It has been developed as an easy-to-use simulation tool with small data requirements. Even though complexity is reduced to the essentially necessary level in favor of applicability, the model is able to represent major land use development, i.e. population and employment, and major activity patterns, i.e. commute trips and goods' flows. Based on its simple assumptions it is transparent and easy to use.

Generally, the model fulfills its purpose. However, the model design has limitations and shortcomings that shall be made explicit. SEAM as a gravity model does not simulate decisions or transitions explicitly and, therefore, is less behaviorally rich than discrete choice models. It does not have the power of economic input-output models as SEAM is not driven by economic development but rather by exogenous forecasts. The model reaches an equilibrium in every simulation period even though there is a broad consensus that this equilibrium is never truly reached in reality, just because households and employers react with a time lag on changes in the urban environment and make decisions based on perceived utility that is biased by preference, prejudice and limited information. There is a tradeoff between complexity and easiness to implement a model. In this case, simplicity was chosen to ensure a robust and easy to use simulation tool.

While the validation results are convincing for population, simulation of employment lacks accuracy in certain industries. This result is little surprising. Many advanced urban simulation models perform very well for population and reach much lower correlations with real world development on the employment side, even after years of development. This is due to several reasons, including a larger heterogeneity of businesses compared to population, lack of theory and data about business development, and the impact of unpredictable historic events such as closing down an entire large plant. Labor flows and commodity flows appear to be represented well as they match quite closely the available target data. However, it is advisable to be cautious to evaluate the activity generation based on this short validation. The overall aggregate flows indicate that real flows are resembled, but further analysis would be desirable to check the spatial distribution of flows.

This relatively simple model structure and straightforward calibration approach has been shown to produce promising results in a very short time period. SEAM within the OSMP Model shows valuable capabilities for policy scenarios encouraging further development as well as the application in additional in other study areas. It must also be noted, however, that the rapid deployment of SEAM would not have been possible without the availability of the base data, selected model outputs and the forecast control totals provided by the OSMP project. Overall, SEAM can be seen as a valid land use and activity simulation capable to improve transportation models to integrated land-use/transportation models.

7 References

Arentze, Theo and Harry Timmermans (2000) *ALBATROSS - A Learning Based Transportation Oriented Simulation System*. Eindhoven: European Institute of Retailing and Services Studies.

de la Barra, Tomás (1989) *Integrated Land Use And Transport Modelling. Decision Chains and Hierarchies*. Cambridge Urban And Architectural Studies 12. Cambridge: Cambridge University Press.

de la Barra, Tomás, B. Pérez and N. Vera (1984) TRANUS-J: putting large models into small computers. *Environment and Planning B: Planning and Design* 11: 87-101.

Deming, W. Edwards and Frederick F. Stephan (1940) On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known. *The Annals of Mathematical Statistics* 11: 427-444.

Echenique, Marcial H., A.D.J. Flowerdew, John D. Hunt, T.R. Mayo, I.J. Skidmore, David C. Simmonds (1990) The MEPLAN models of Bilbao, Leeds and Dortmund. *Transport Reviews* Vol. 10 Iss. 4: 309-322

Echenique, Marcial H., D. Crowther, W. Lindsay (1969) A Spatial Model of Urban Stock and Activity. *Regional Studies* 3: 281-312.

Forrester, Jay W. (1969) *Urban Dynamics*. Cambridge (Massachusetts): The M.I.T. Press.

Goldner, William, Stephen R. Rosenthal and Jack R. Meredith (1972) *Theory and Application: Projective Land Use Model 2*. Berkeley: Institute of Transportation and Traffic Engineering.

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

Harris, Britton J. (1966) *Basic Assumptions for a Simulation of the Urban Residential Housing and Land Market*. Working Paper. Philadelphia: University of Pennsylvania.

Herbert, John D. and Benjamin H. Stevens (1960) A model for the distribution of residential activity in urban areas. *Journal of Regional Science* 2: 21-36.

Hunt, John D. and John E. Abraham (2003) Design and Application of the PECAS Land Use Modelling System. *Proceedings of the 8th Conference on Computers in Urban Planning and Urban Management*. Sendai, Japan.

Hunt, John D., David S. Kriger and Eric J. Miller (2005) Current Operational Urban Land-use-Transport Modelling Frameworks: A Review. *Transport Reviews* 25: 329-376.

Kain, John F. (1987) Computer simulation models of urban location. In: E. S. Mills (Ed.) *Handbook of regional and urban economics*. Volume II: Urban Economics, vol. 7, Handbook in Economics. Edited by K. J. Arrow and M. D. Intriligator. Amsterdam: North-Holland. 847-875.

Kanaroglou, Pavlos S. and Darren M. Scott (2002) Integrated Urban Transportation and Landuse Models for Policy Analysis. In: M. Dijst, W. Schenkel and I. Thomas (Eds.) *Governing Cities on the Move. Functional and management perspectives on transformations of European urban infrastructures*, Urban and Regional Planning and Development. Hamshire (England), Burlington (VT): Ashgate. 42-72.

Kreibich, Volker (1978) The successful transportation system and the regional planning problem: an evaluation of the Munich rapid transit system in the context of urban and regional planning policy. *Transportation* 7: 137-145.

Landis, J. and M. Zhang (1998a) The second generation of the California urban futures model. Part 1: Model logic and theory. *Environment and Planning B: Planning and Design* 25: 657-666.

Landis, J. and M. Zhang (1998b) The second generation of the California urban futures model. Part 2: Specification and calibration results of the land-use change submodel. *Environment and Planning B: Planning and Design* 25: 795-824.

Lowry, Ira S. (1964) *A Model of Metropolis*. Memorandum RM-4035-RC. Santa Monica: Rand Corporation.

Martínez, Francisco J. (1996) MUSSA: Land Use Model for Santiago City. *Transportation Research Record* 1552: 126-134.

Moeckel, Rolf, Björn Schwarze, Klaus Spiekermann and Michael Wegener (2007) Simulating Interactions between Land Use, Transport and Environment. *Proceedings of the 11th World Conference of Transport Research*. Berkeley, California.

Putman, Stephen H. (1983) Integrated Urban Models. Policy Analysis of Transportation And Land Use. Research in Planning and Design 10. London: Pion.

Reynolds, Marilyn M. and Jack R. Meredith (1972) *Computer Systems Guide: Projective Land Use Model 3*. Berkeley: Institute of Transportation and Traffic Engineering.

Rosenthal, Stephen R., Jack R. Meredith and William Goldner (1972) *Plan making with a computer model: Projective land use model 1*. Berkeley: Institute of Transportation and Traffic Engineering.

Schürmann, Carsten, Klaus Spiekermann and Michael Wegener (1997) *Accessibility indicators*. Berichte aus dem Institut für Raumplanung 102: University of Dortmund.

Simmonds, David C. (1999) The design of the DELTA land-use modelling package. *Environment and Planning B: Planning and Design* 26: 665-684.

Simmonds, David C. (2001) The Objectives and Design of a New Land-use Modelling Package: DELTA. In: G. Clarke and M. Madden (Eds.) *Regional Science in Business, Advances in Spatial Sciences*. Berlin: Springer. 159-188.

Strauch, Dirk, Rolf Moeckel, Michael Wegener, Jürgen Gräfe, Heike Mühlhans, Guido Rindsfuser and Klaus J. Beckmann (2005) Linking Transport and Land Use Planning: The Microscopic Dynamic Simulation Model ILUMASS. In: P. M. Atkinson, G. M. Foody, S. E. Darby and F. Wu (Eds.) *GeoDynamics*. Boca Raton: CRC Press. 295-311.

Timmermans, Harry (2003) The Saga of Integrated Land Use-Transport Modeling: How Many More Dreams Before We Wake Up? Conference keynote paper. *Proceedings of Moving through nets: The physical and social dimensions of travel. 10th International Conference on Travel Behaviour Research*. Lucerne. 35.

U.S. Environmental Protection Agency (2000) *Projecting Land-Use Change. A Summary of Models for Assessing the Effects of Community Growth and Change on Land-Use Patterns*. Cincinnati: U.S. Environmental Protection Agency, Office of Research and Development.

Waddell, Paul (2002) UrbanSim. Modeling Urban Development for Land Use, Transportation, and Environmental Planning. *Journal of the American Planning Association* 68: 297-314.

Waddell, Paul, Alan Borning, Michael Noth, Nathan Freier, Michael Becke and Gudmundur F. Ulfarsson (2003) Microsimulation of Urban Development and Location Choice: Design and Implementation of UrbanSim. *Networks and Spatial Economics* 3: 43-67.

Wegener, Michael (1982) Modeling Urban Decline: A Multilevel Economic-Demographic Model for the Dortmund Region. *International Regional Science Review* 7: 217-241.

Wegener, Michael (1994) Operational Urban Models. *Journal of the American Planning Association* 60: 17-29.

Wegener, Michael (1998a) *The IRPUD Model: Overview*. Internet Resource: http://www.raumplanung.uni-dortmund.de/irpud/pro/mod/mod_e.htm. Accessed 30 July 2007.

Wegener, Michael (1998b) Applied Models of Urban Land Use, Transport and Environment: State of the Art and Future Developments. In: L. Lundqvist, L.-G. Mattsson and T. J. Kim (Eds.) *Network Infrastructure and the Urban Environment. Advances in Spatial Systems Modelling*, Advances in Spatial Science. Berlin: Springer. 245-267.

Wegener, Michael (2004) Overview of Land use Transport Models. In: D. A. Hensher, K. J. Button, K. E. Haynes and P. R. Stopher (Eds.) *Handbook of Transport Geography And Spatial Systems*, vol. 5, Handbooks in Transport. Edited by D. A. Hensher and K. J. Button. Amsterdam: Elsevier. 127-146.