A Framework for Visualizing Multivariate Geodata

Ariane Middel
Computer Science Department
University of Kaiserslautern

Abstract: In urban planning, sophisticated simulation models are key tools to estimate future population growth for measuring the impact of planning decisions on urban developments and the environment. Simulated population projections usually result in bulky, large-scale, multivariate geospatial data sets. Millions of records have to be processed, stored, and visualized to help planners explore and analyze complex population patterns.
This paper introduces a database driven framework for visualizing geospatial multivariate simulation data from *UrbanSim*, a software-based simulation model for the analysis and planning of urban developments. The designed framework is extendable and aims at integrating methods from information visualization and cartography into planning processes.

1 Introduction

Estimates indicate that nowadays, approximately 80 percent of digital data is geospatially referenced [Nat03]. An ever-increasing availability of geodata raises the demand for new visual representations which are beyond classical 2D maps, particularly with regard to increasing data set size and dimensionality. This applies for acquirable real-world data as well as for estimated and projected data.
In urban planning, simulation models for predicting population growth have become highly complex. Model calculations result in large-scale, multidimensional spatial data sets. Projection data needs to be managed, manipulated, and visualized in order to enable planners to visually compare different planning scenarios and to evaluate simulated impacts of different land use policies. Especially the visualization task remains challenging, since traditional map-centered approaches lack support for high multivariability. The emerging research field Geovisualization (GeoVIS) takes advantage of techniques from cartography, geographic information science, and computer science to overcome visualization issues dealing with multi-attributed, large-scale spatial datasets.
Within the scope of this research domain, we develop a database driven framework to assist planners in their geovisual analyses of multidimensional simulation data. The framework uses *UrbanSim* demographic projection data for Maricopa County on a households-per-grid-cell basis. Additional residential building type data is derived from demographic household characteristics by statistical regression analysis. A geodatabase stores both simulated and estimated data for subsequent visualization. Within the database, modular scripts generate georeferenced scalable geometries calculated on the basis of multiple data attribute values for a map-based geovisualization.

2 Related Work

Visualizing geospatial datasets has long been a key issue in cartography. The cartographer Bertin established a basis for designing maps in his classical work "Semiology of Graphics" [Ber67] where he identified a set of fundamental visual variables and defined graphical rules for their appropriate use. Since then, Bertin's concepts have been constantly modified and extended. Modern cartography transfers design knowledge from 2D paper maps to new media. On-screen interactive maps are designed to assist in visual data exploration and analyses [AA99]. Cartographic visualization is also extended to abstract and non-geographic data by spatialization [SF03].

Whereas cartography primarily deals with representations constrained to a spatial domain, information visualization (InfoVIS) is mainly concerned with the display of large multivariate datasets. In the early 70's, Chernoff presented a technique to visualize trends in highly dimensional data by relating data to facial features [CR75]. Gradually over the years, new information visualization techniques were introduced, ranging from 2D scattergraphs to 3D treemaps. For a comprehensive overview of developments in Information visualization we refer to [SCM99] and [Tuf90].

Recently, efforts have emerged to combine techniques from both cartography and information visualization [Sku00, FS04]. Geographic visualization (GeoVIS) is a new, rapidly evolving domain, especially since the availability of geodata is increasing. In 1998, MacEachren compiled a first research agenda entitled "Visualization - Cartography for the 21st century" [Mac98] and addressed GeoVIS research challenges. Since then, cartographic and InfoVIS techniques have been applied to design integrated geovisualization tools frequently. Latest advances include multivariate analyses with self-organizing maps [GGMZ05, SH03], studies on human activity patterns using 3D space-time paths [MPJ04], and bivariate maps for public health studies [MGP$^+$04]. Most recent activities in geovisualization research are discussed in [Kra06].

Pinnel et al. [PDBB00] conducted a study on visualization designs for urban modeling. They found out that map-centered visualizations are the most useful portrayals for urban planning and analysis, since map layout encodes location information which is crucial for decision-making. A map-based visualization approach, "The Indicator Browser", was designed by Schwartzman et al. [SB07] to display *UrbanSim* simulation results. The browser uses comparative visualizations of 2D maps to satisfy multivariability which impedes human vision to recognize complex patterns across many dimensions.

Instead of encoding n-dimensional data in n maps, Tominski et al. [TSWS05] follow a different approach to display monthly health data. They visualize time dependent multivariate disease information as 3D pencil and helix icons geocoded on a 2D map. The research reported here helps to analyze complex patterns across multivariate, spatial, and temporal dimensions, but it lacks a powerful geodatabase and GIS functionality to manage, process, and distribute data. In the following, we will present a geovisualization framework to overcome those drawbacks.

3 A Framework for Visualizing Multivariate Geodata

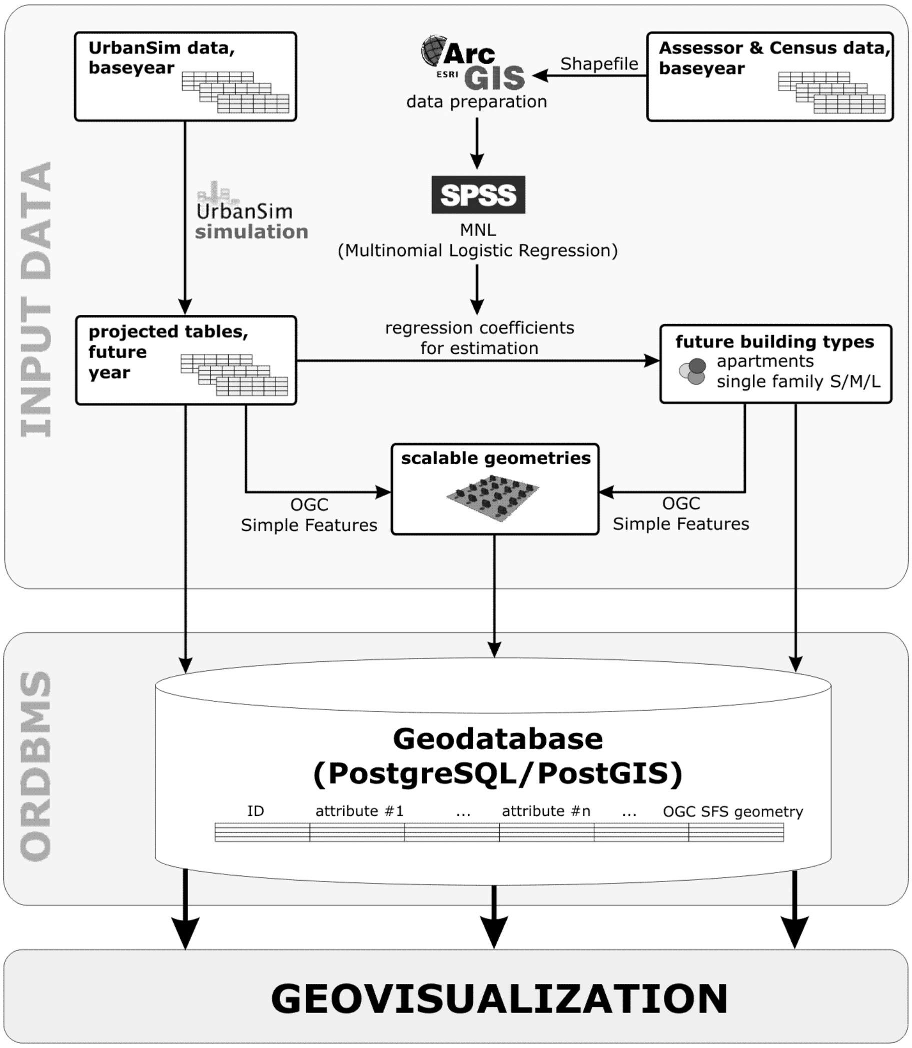

Figure 1: System Architecture (reproduced in color on p. 11)

The geovisualization framework mainly consists of three parts (cp. Fig. 1): an input data layer, an object-relational data base management system, and a visualization layer. Within the data layer, demographic data is aggregated from different sources and prepared in ArcGIS. An *UrbanSim* simulation projects demographic household characteristics for a predefined year in the future (see Section 3.1). For the same year, demographic data is aggregated to estimate future residential building types with multinomial logistic regression in SPSS (see Section 3.2). The demographic simulation results and the estimated building types are projected into 3D space with scalable, georeferenced geometries (see Section 3.4). Data and geometry is stored and managed in an object-relational geodatabase (see Section 3.3) and finally visualized on top of map a. In the remainder of the paper, we will explain the system architecture in detail.

3.1 Data modeling with *UrbanSim*

UrbanSim [BW04] is a large scale land use and transportation simulation software to model the possible long-term effects of different policies on urban developments. More precisely, it simulates the interactions between transportation, land use, and public policy at household and job level. *UrbanSim* consists of numerous model components simulating different actors in the urban development process, e.g. discrete choice models for relocating households and jobs. The open source simulation model was developed by a research group in Washington and is currently implemented in the Digital Phoenix project [Sub07] at ASU, Arizona State University to predict population growth in the Phoenix Metropolitan area.

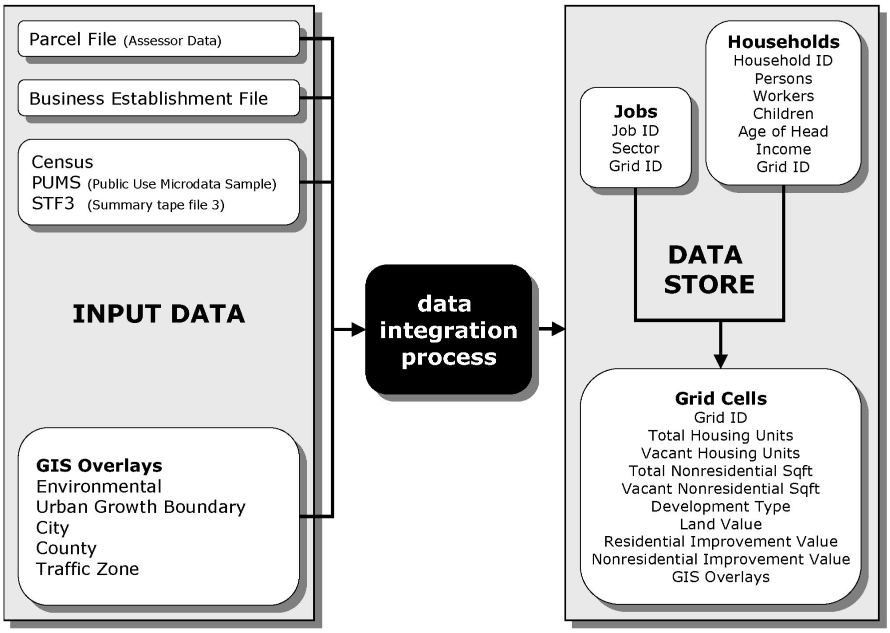

Figure 2: Data integration process (c.f. [Wad02])

UrbanSim input data is aggregated from various sources (see Fig. 2) and spatially mapped to a georeferenced grid cells file. Typical grid cell sizes are $150m \times 150m$ or 1 square

mile. Base year data includes information on parcels from the Assessor's office, employment data, and Census data. Additional input data on city, county, and urban growth boundaries as well as environmental and traffic information is overlaid in ArcGIS. The *UrbanSim* data store contains a grid cells table, a jobs table holding information on each job and its employment sector in the grid cells, and a household table. The latter is synthesized probabilistically and compiles demographic characteristics for each household in the metropolitan area.
Subsequent to the data integration process, the *UrbanSim* simulation is run for a predefined number of years. The output projection tables include data on future households with grid cell location and demographic characteristics, future jobs, and exogenous input data. The projection results can be integrated into a variety of analyses, e.g. the analysis of future population density, material use, or carbon footprints.

3.2 Estimating Residential Bulding Types from Demographic Data

Knowledge of future residential building types is essential for the above mentioned analyses and for visualization purposes. To derive the type of residential units from *UrbanSim* demographic data for each grid cell we assume a relationship between household characteristics and building types. Multinomial Logistic regression (MNL) is an analytically appropriate technique to link demographic attributes with dwelling types. This statistical method is widely used in social sciences and economics to discover hidden relationships between variables. Multinomial logistic regression is equivalent to the conditional discrete choice model, first presented and most notably influenced by McFadden [McF73, McF76, McF78]. MNL models the interaction between response and explanatory variables. The multinomial response is polytomous, it can have multiple unordered qualitative categories as outcomes. The parameter values of the regression are estimated according to the following equation [PX99]:

$$z_{ij} = \boldsymbol{x}_i' \boldsymbol{\beta}_j = \sum_{k=0}^{K} \beta_{jk} x_{ik} = \alpha_j + \sum_{k=1}^{K} \beta_{jk} x_{ik} \tag{1}$$

Here, $\boldsymbol{x}$ is the vector of predictors storing all relevant demographic variables and $\boldsymbol{\beta}$ denotes the regression parameter vector containing estimation coefficients.
With the calculated set of coefficients β_{jk}, the probability P_{ij} that a certain building type category j is chosen over any other building type category yields:

$$P(y_i = j | \boldsymbol{x}_i) = P_{ij} = \frac{e^{z_{ij}}}{\sum_{j=1}^{J} e^{z_{ij}}} = \frac{e^{\boldsymbol{x}_i' \boldsymbol{\beta}_j}}{\sum_{j=1}^{J} e^{\boldsymbol{x}_i' \boldsymbol{\beta}_j}} = \frac{e^{\boldsymbol{x}_i' \boldsymbol{\beta}_j}}{1 + \sum_{j=2}^{J} e^{\boldsymbol{x}_i' \boldsymbol{\beta}_j}} \tag{2}$$

Basically, we distinguish between multifamily dwellings and single family dwellings. Predictable categories for the regression analysis are apartments as well as single family houses with small, medium, and large lots. Those categories can be further refined if additional data is available, e.g. the number of storeys.

Variables explaining the building types are primary demographic characteristics corresponding to the *UrbanSim* output: household income, average household size, number of households, median age, presence of children, and percentage of minorities and Hispanics. Demographic data is compiled and synthesized to the *UrbanSim* grid cell base file from Assessor's data and Census block groups. Afterwards, the regression analysis is run in SPSS with the presented explanatory variables and building type categories. To predict future building types, we apply the estimated coefficients β_{jk} to *UrbanSim* simulation data. The resulting dwelling categories are stored together with demographic attributes from *UrbanSim* in a geodatabase, as explained in the next section.

3.3 Geodatabase

The inherent geospatial nature of *UrbanSim* gridded data requires the implementation of a spatial database. So-called geodatabases extend the database concept to storage, query, and editing of georeferenced objects. Accordingly, the projected demographic data from *UrbanSim* and the estimated building type data is stored in a geodatabase implemented in PostgreSQL [Pos07], an open source object-relational database management system (ORDBMS). In addition, a PostGIS module [SLRL05] serves as geospatial extension to the PostgreSQL backend server. PostGIS enables PostgreSQL to integrate spatial data structures in the database, query geographic objects, and serve them to GIS applications.
In general, each spatial table in the geodatabase represents a separate PostGIS layer. An ancillary table contains meta-data on the associated geodetic datum, here the coordinate reference system NAD 1983. Each distinct geographic object constitutes a record in a spatial table and associated attribute information is stored in data columns. In particular, the implemented geodatabase includes a table with records for *UrbanSim* grid cells and columns for *UrbanSim* projected demographic variables and the associated estimated building type category.
PostGIS provides a dedicated geometry column which contains geometric information for each feature in the form of point, line, or polygon data types. We will use this geometry column to store the scalable geometries generated from attribute data for the geovisualization.

3.4 Geometry Generation and Visualization

The designed PostgreSQL geodatabase provides the basis for a multivariate information visualization. *UrbanSim* simulation data and estimated building types stored in the database are encoded in scalable 3D geometries and visualized in a geospatial context on top of a map. Geometric objects are created for every grid cell directly within the geodatabase from geometric primitives.
PostGIS is fully OGC compliant which means it conforms to the "Simple Features for SQL" specification (SFS) from the OpenGIS Consortium [Ope07]. The OGC SFS defines standard geometric data types (see Fig. 3) and functions for manipulating geometry.

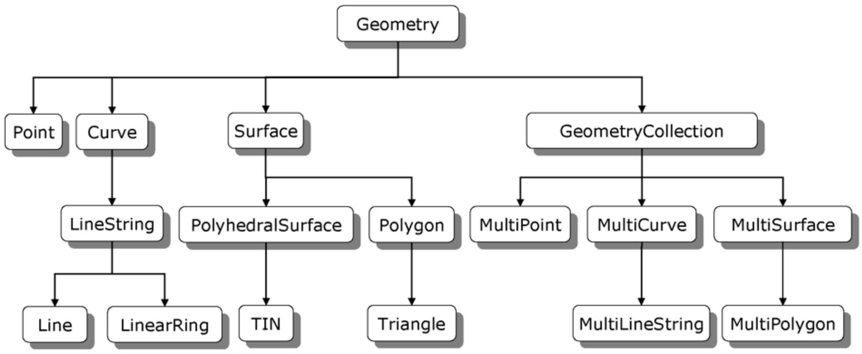

Figure 3: OGC Simple Feature Specification (cf. [Ope07])

PostGIS supports 1D and 2D geometric primitives but lacks geometric data types for volumes. Therefore, 3D objects have to be constructed using 3D polygons. Geometry calculations are performed using dedicated PostGIS geometry functions. For each record in the geodatabase, modular SQL scripts select the associated attributes to be visualized and generate user-specified geometries. At the same time, size and shape of the geometries are scaled according to normalized attribute values.
The geometry generating SQL modules can be implemented to calculate arbitrarily shaped discrete geometries. To visualize urban structures and household demographics, an implementation of iconized buildings seems obvious. In this case, shape parameters to vary by attribute values can be floor plan, roof height, ridge height or the number of chimneys for instance. Other possible geometries include stepped pyramids, discretized cones and tori. The geometric objects are generated for each *UrbanSim* grid cell and stored in the geodatabase. From there, the geometries can be geovisualized in a map context, e.g. with Google Earth.

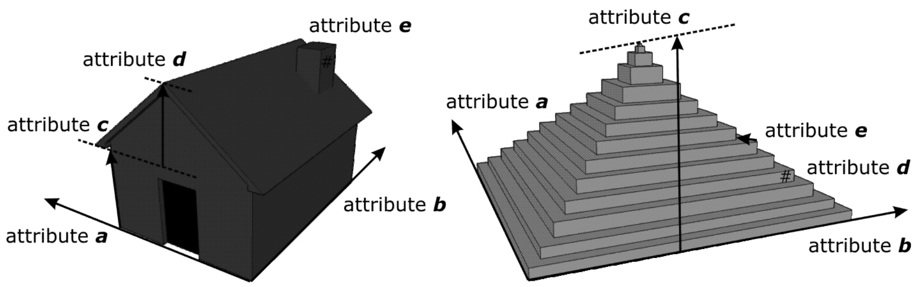

Figure 4: Examples for scalable geometries

4 Conclusions and Future Work

This paper presented an integrated framework for processing, storing, and visualizing multivariate geodata. The introduced framework uses simulation data from *UrbanSim*, a planning tool for the simulation, comparison, and evaluation of different planning scenarios to better assess the impacts of various policy decisions. Within the framework, residential building types are estimated from demographic household characteristics by means of multinomial logistic regression. The aggregated output data is stored in a PostgreSQL geodatabase and provides the basis for further visualization.

The developed visualization approach is a cross-disciplinary effort to integrate methods from InfoVIS and cartography for supporting space-related decision-making. Geovisualization is driven by the need to visualize geospatial data which is multivariate, large, and multidimensional at the same time. Our framework uses a map-centered visualization where data attributes are represented by georeferenced scalable 3D geometries. Geometric shapes are generated within the geodatabase and scaled according to normalized attribute values.

The presented geovisualization approach is extendable, since data representation is not restricted to specified geometric shapes. SQL modules can easily be added to generate various geometries, e.g. cones or tori instead of iconic dwellings. This makes the framework generic enough to be transferable to other application areas dealing with any kind of multidimensional geodata.

Besides extending the variety of geometry modules, future work includes the implementation of an interactive web-based 3D visualization. A PHP middleware will be designed, allowing the user to interactively choose the parameters for the geometries and the attributes he wishes to visualize from the geodatabase. A KMZ file will be generated from the database according to the user's geometry specifications and visualized in Google Earth on the fly. An internet based choice and display of data will provide the framework to a variety of user groups and increase the number of geovisualization tasks to be solved.

Acknowledgements

This work was supported by the German Science Foundation (DFG, grant number 1131) as part of the International Graduate School (IRTG) in Kaiserslautern on "Visualization of Large and Unstructured Data Sets. Applications in Geospatial Planning, Modelling, and Engineering".

References

[AA99] Gennady L. Andrienko and Natalia V. Andrienko. Interactive Maps for Visual Data Exploration. In *International Journal Geographic Information Science*, volume 13, pages 355–374, June 1999.

[Ber67] Jacques Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1967.

[BW04] Alan Borning and Paul Waddell. Integrated Land Use, Transportation, and Environmental Simulation: UrbanSim Project Highlights. In *dg.o '04: Proceedings of the 2004 annual national conference on Digital government research*, pages 1–2. Digital Government Research Center, 2004.

[CR75] H. Chernoff and M. H. Rizvi. Effect on Classification Error or Random Permutations of Features in Representing Multivariate Data by Faces. *Journal of American Statistical Association*, 70:548–554, 1975.

[FS04] Georg Fuchs and Heidrun Schumann. Visualizing Abstract Data on Maps. *IV '00*, 00:139–144, 2004.

[GGMZ05] Diansheng Guo, Mark Gahegan, Alan M. MacEachren, and Biliang Zhou. Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach. *Cartography and Geographic Information Science*, 32(2):113–132, April 2005.

[Kra06] Menno-Jan Kraak. Visualization Viewpoints: Beyond Geovisualization. *IEEE Computer Graphics and Applications*, 26(4):6–9, Jul/Aug 2006.

[Mac98] Alan M. MacEachren. Visualization - Cartography for the 21st century. In *Proceedings of the 7th Annual Conference of Polish Spatial Information Assocaition*, Warsaw, Poland, May 1998.

[McF73] Daniel L. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*, pages 105–142, 1973.

[McF76] Daniel L. McFadden. *Properties of the Multinomial Logit (MNL) Model*. Urban Travel Demand Forecasting Project, Institute of Transportation Studies, 1976.

[McF78] Daniel L. McFadden. *Modelling the Choice of Residential Location*. Institute of Transportation Studies, University of California, 1978.

[MGP+04] Alan M. MacEachren, Mark Gahegan, William Pike, Isaac Brewer, Guoray Cai, Eugene Lengerich, and Frank Hardisty. Geovisualization for Knowledge Construction and Decision Support. *IEEE Computer Graphics and Applications*, 24(1):13–17, Jan/Feb 2004.

[MPJ04] Kwan Mei-Po and Lee Jiyeong. Geovisualization of Human Activity Patterns Using 3D GIS: A Time-Geographic Approach. In Michael F. Goodchild M and Donald G. Janelle, editors, *Spatially Integrated Social Science: Examples in Best Practice*, pages 48–66, Oxford, 2004. Oxford University Press.

[Nat03] National Research Council. *IT Roadmap to a Geospatial Future*. National Academies Press, 2003.

[Ope07] Open Geospatial Consortium. OGC Simple Feature Specification. Available online at: <http://www.opengeospatial.org/standards>, 2007. (Last Accessed: 09-23-2007).

[PDBB00] L. Denise Pinnel, Matthew Dockrey, A. J. Bernheim Brush, and Alan Borning. Design of Visualizations for Urban Modeling. In *VisSym '00: Joint Eurographics – IEEE TCVC Symposium on Visualization*, Amsterdam, The Netherlands, May 2000.

[Pos07] PostgreSQL Global Development Group. PostgreSQL. Available online at: <http://www.postgresql.org>, 2007. (Last Accessed: 09-23-2007).

[PX99] Daniel Powers and Yu Xie. *Statistical Methods For Categorical Data Analysis*. Elsevier Science & Technology, 1999.

[SB07] Yael Schwartzman and Alan Borning. The Indicator Browser: A Web-Based Interface for Visualizing UrbanSim Simulation Results. In *HICSS '07: Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, page 92, Washington, DC, USA, January 2007. IEEE Computer Society.

[SCM99] Ben Shneiderman, Stuart K. Card, and Jock D. Mackinlay. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, CA, 1999.

[SF03] André Skupin and Sara I. Fabrikant. Spatialization Methods: A Cartographic Research Agenda for Non-Geographic Information Visualization. In *Cartography and Geographic Information Science*, volume 30, pages 95–115, 2003.

[SH03] André Skupin and Ron Hagelman. Attribute Space Visualization of Demographic Change. In *11th ACM GIS Symposium (GIS'03)*, New Orleans, Louisiana, USA, November 7-8 2003. ACM Press.

[Sku00] André Skupin. From Metaphor to Method: Cartographic Perspectives on Information Visualization. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVIS 2000)*, pages 91–97, Salt Lake City, UT, USA, October 9-10 2000. Los Alamitos: IEEE Computer Society.

[SLRL05] Sandro Santilli, Mark Leslie, Paul Ramsey, and Jeff Lounsbury. PostGIS. Available online at: <http://postgis.refractions.net>, 2005. (Last Accessed: 09-22-2007).

[Sub07] Subhrajit Guhathakurta and Janet Holston and Yoshi Kobayashi and Tim Lant and Mookesh Patel. Digital Phoenix Project. Available online at: <http://www.digitalphoenix-asu.net>, 2007. (Last Accessed: 09-26-2007).

[TSWS05] Christian Tominski, Petra Schulze-Wollgast, and Heidrun Schumann. 3D Information Visualization for Time Dependent Data on Maps. In *IV '05: Proceedings of the Ninth International Conference on Information Visualisation (IV'05)*, pages 175–181, Washington, DC, USA, 2005. IEEE Computer Society.

[Tuf90] Edward Tufte. *Envisioning Information*. Graphics Press, 1990.

[Wad02] Paul Waddell. UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning. *Journal of the American Planning Association*, 68(3):297–314, 2002.

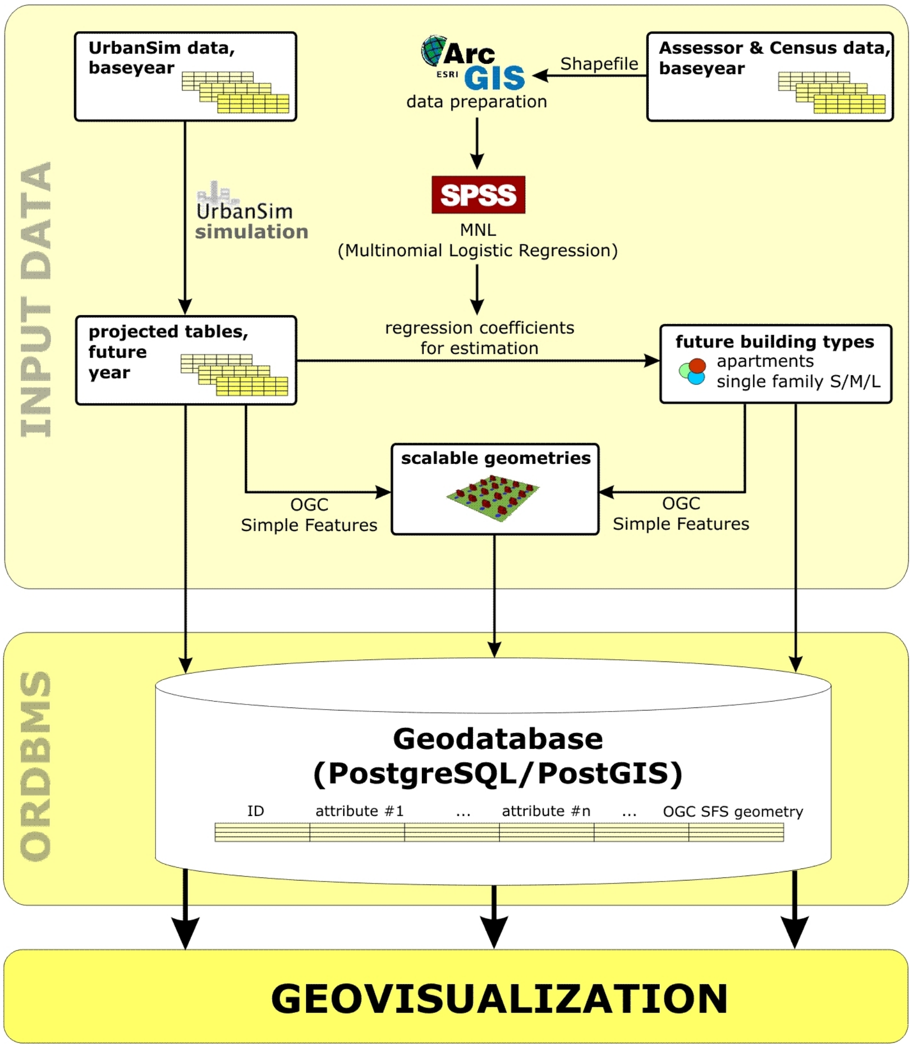

Figure 5: System Architecture