The *luci2 Urban Simulation Model* and the Central Indiana Implementation

Working Paper

Version 1.0
March 20, 2007

John R. Ottensmann

jottensm@iupui.edu

Center for Urban Policy and the Environment
Indiana University-Purdue University Indianapolis
342 N. Senate Avenue
Indianapolis, IN 46202

Contents

Introduction

The *luci2 Urban Simulation Model* is a general-purpose model to simulate future urban development. The model is initially being implemented for a 44-county region in Central Indiana. This paper documents the development of the *luci2 Urban Simulation Model* and the Central Indiana Implementation.

The *luci2 Urban Simulation Model* is the successor to LUCI: Land Use in Central Indiana Model. (Information on the LUCI model and the model itself can be obtained from the model website at http://luci.urbancenter.iupui.edu/.) The new model extends the original model in several important ways: *luci2* separately simulates residential and employment-related development. The model forecasts employment growth by ZIP code for four industry sectors. And the *luci2* program is being developed so similar models, of varying levels of complexity, can readily be implemented for other areas.

The *luci2 Model* predicts the general trends in urban development over an extended period. It cannot and will not provide detailed, specific predictions of which parcels will develop. Rather, by simulating the general pattern of future development, the model enables users to examine different policy options and assumptions to determine their effects on the general pattern of development.

The model is intended to be used to produce different development scenarios reflecting policy choices and alternative assumptions regarding future patterns of development that can be considered for the region. Examples of such scenarios and policy choices include restricting development in along streams and in wetlands, policies to preserve agricultural land, new transportation developments, and increased densities of development. The model produces maps showing expected future patterns of development and measures associated with each scenario such as the length of the journey to work and amounts of different types of land converted to urban use.

Like its predecessor, the *luci2* model has been designed so persons with no expertise can use the model to create and examine scenarios. Also, when completed, the Central Indiana Implementation will likewise be freely distributed. (The general *luci2 Urban Simulation Model* with the capabilities to create new implementations for other areas and with extended capabilities will not be freely distributed.)

A note on the name: The *luci2 Urban Simulation Model* is the successor to LUCI, Land Use in Central Indiana Model. As such, the desire was to continue the use of the "*luci*" name. However, the *luci2 Urban Simulation Model* is a general purpose model not tied to Central Indiana, and "*luci*" would not longer be an acronym. Using the name "*luci2*" in all lower-case seemed to best distinguish the new model from its predecessor while retaining the name.

Model Structure

The model simulates the conversion of land from nonurban uses to residential uses and employment-related uses for a system of simulation zones, 1-mile-square grid cells in the Central Indiana Implementation. The simulation is driven by an exogenous forecast of population growth for the entire region and proceeds in 5-year simulation periods in this implementation.

The model can best be characterized as using random utility theory and aggregated discrete choice models of the decisions to convert available nonurban land to residential and employment-related uses. The model predicts the choice process of producers leading to the conversion of land from vacant to residential and employment-related uses. The assumption is that the producers incorporate in their decisions their assessments of the demand by consumers for residential land use and firms for employment-related land use. The model also predicts population densities for new residential development and the employment densities for new employment-related development.

The predictive models do not consider individual choices but rather the aggregation of individual choices at the level of the simulation zones. Using aggregated data, aggregate logit models predict the probabilities of decisions to convert smaller units of land within the simulation zones from available nonurban uses to residential uses and employment-related uses.

Population growth is assumed to create demand for new residential development. Two models have been estimated for use in simulating that development. The aggregate logit model predicts the probability of available nonurban land in each simulation zone being converted from nonurban to residential use. A second model predicts the population density for each simulation zone. For each 5-year simulation period, the probability of residential development and the population density are predicted for each zone. The predicted probability is multiplied by the amount of available nonurban land in each zone and the predicted density to give an initial value for the population that could be accommodated. Then the probabilities are adjusted across the region to provide the amount of residential development required to accommodate the predicted population growth.

Employment is assumed to grow at the same rate as the population. Change in employment is forecast for employment zones—ZIP codes for the Central Indiana Implementation—for four industry sectors. A model predicts the land area required per employee for each of the employment zones to provide the amount of new employment-related development required for each zone as a result of the forecast employment growth. Finally, new employment-related development is allocated to the simulation zones using another aggregate logit model predicting the probability of available nonurban land being converted to employment-related uses in each simulation zone, along with constraints to achieve reasonable clustering of employment-related development.

Model Capabilities

This section provides an overview of the capabilities of the *luci2 Urban Simulation Model* program. More details on aspects of the model simulation, implementation of the scenarios, and procedures for implementing models are provided in later sections.

Adaptable Program Structure

The *luci2* program is a general-purpose urban simulation program that can be adapted for different areas, model structure, variables, and so forth. It will not be restricted to use only for the Central Indiana Implementation. The capabilities of the program have, however, been designed especially to accommodate the Central Indiana Implementation.

Following are examples of the types of specifications that can be made in implementing a model:

- The number of simulation zones, employment zones, and model subareas
- The use of external zones outside the simulation area for areas with significant external interactions
- Whether the model uses a system of regular grid cells or irregular areas such as Traffic Analysis Zones
- The use of English or metric units
- Whether the model uses distances or travel times in the calculation of accessibility measures
- Whether the model includes the capabilities for predicting employment, either local service employment and/or employment for multiple industry sectors
- Whether the model predicts residential and employment-related development separately or predicts urban development as in LUCI
- The specific variables to be included in the various predictive models and the forms of those variables
- Specific scenario options to be included in the model

The program can be used to implement models for different areas and datasets without requiring modification of the program. Such implementations will not be trivial. The implementation will involve assembling the data to be used, estimating the equations, creating datasets in the required formats, and creating the model specification file. Still, the implementation of a model for a different area will be considerably easier than if it were necessary to modify or rewrite the program.

The program can be used to implement models of widely varying complexity. The simplest model would be a model in which employment accessibility or some other variable would be used to predict the conversion of land to urban use. The most complex model would be a model with the complexity of the Central Indiana Implementation. The original LUCI model could actually be implemented using the *luci2* model. LUCI/T, the custom adaptation for the Central Indiana Suburban Transportation & Mobility Study, could also be implemented using *luci2*. Statewide versions of *luci2* could be implemented with varying levels of complexity.

Model Scenarios

The program provides options for implementing a wide range of scenario options, including all of those included in the original LUCI model along with several others. Following are the scenario possibilities:

- Target year
- Population growth, including user-specified rates of growth, external population forecasts, and population forecasts for subareas that can be used as control totals
- Population densities higher or lower than predicted by model or minimum or maximum densities
- Expansion of utilities as development occurs, user-specified expansion of utilities, options to require utilities, option to require sewer utility service if land is not suitable for septic systems (the last is not included in the Central Indiana Implementation
- Planned transportation improvements
- Restriction of development on various types of sensitive lands
- Restriction of development on agricultural land
- Restriction that development occur in accord with comprehensive plan or zoning restrictions (not included in the Central Indiana Implementation)
- Restriction of development to within user-specified urban growth boundaries
- User-specification that development be more or less dispersed than otherwise predicted by model
- Relative importance of accessibility to employment
- Method of allocating employment growth—local-service and basic employment predicted by model; local-service predicted by model and basic allocated using external forecast or simpler allocation methods; all employment allocated using simpler allocation methods; external forecast of employment
- Addition of new employment at specified location
- Specification that subarea grows more or less rapidly than otherwise predicted by model
- Reduction in differences in other variable used for predicting residential development
- Advanced options including changes to internal demand for new development, maximum development in a simulation zone, minimum predicted probability for new development, and criteria for the allocation of employment-related development

Enhanced User Capabilities

The *luci2* model includes options for user-specified scenario options and for enhanced output that can be used in applying the model for specific applications. These capabilities will allow the use of an implementation of the model to address a broader range of problems without requiring modification to the implementation. For example, the Central Indiana Implementation could be used for applications requiring the definition of scenarios not currently included in the model. Note, however, that these capabilities will not be available with the freely-distributed version of the Central Indiana Implementation.

Users will be able to add scenario options to a model by specifying the scenarios in a user scenario definition file and by providing the data required for the scenario definition in the required format. The types of user-specified scenarios possible will include the following:

- Restriction of development on sensitive land
- Specification of areas in which residential development are allowed
- Specification of minimum densities for residential development
- Specification of maximum densities for residential development
- Specification of areas in which employment-related development are allowed
- Specification of sewer service areas
- Specification of water service areas
- Specification of new transportation improvement

The enhanced user capabilities will also include options to output simulation results beyond those provided in the export results function. These would include the output of more simulation results and the output of these results automatically, either for each simulation period or just for the end of the simulation.

User Interface

The user interface for the *luci2* program has been completely redone from the earlier version. The new interface keeps the approach of placing all of the user options and commands on buttons displayed for the user. This approach was originally taken (and retained) as it is believed that the novel nature of the program would make it hard for a user to know what to do next if the options were hidden away on traditional pull-down menus.

The first improvement is that the Main Menu buttons are simplified and are always displayed and available to the user across the top of the screen. (In the original LUCI program, these buttons were replaced by submenus of buttons when the user made a main menu selection.) For the Tables, New Scenario, and Tools options, the submenus of buttons are now displayed in a column at the left side of the screen. The Maps submenu appears across the top under the Main Menu to retain maximum space for the side-by-side display of two maps.

The selected buttons are highlighted by having their backgrounds changed to white and their captions displayed in bold. This provides the user with clear, consistent cues as to what action is currently being carried out.

Forms for the entry of information for various New Scenario and Tools buttons now appear on the screen directly opposite the button selected. This is a quicker, cleaner procedure than the use of separate dialog boxes in LUCI. (The common dialog for file saving, loading, and printing is retained to provide all of those options in a manner consistent with Windows user-interface conventions.)

The display of information in the tables has been made clearer by presenting scenario results in multiple tables, rather than in the single, very long table that the user was forced to scroll through in the original LUCI model. Key relevant scenario option settings are displayed at the top of the results tables. The table displaying the full set of scenario options has likewise been made more usable by displaying information only on the options selected. With few options selected, the table is short and can be viewed in its entirety without scrolling. Selections of many options, however, can make the table much longer.

Screen shots showing the *luci2* user interface for the major program modes are included in the Appendix.

Inferring Land Use from Remotely Sensed Data

A procedure has been developed to infer residential and employment-related land use from classified land cover data.

The basic data used for this were datasets of classified land cover data derived from multiple LandSat Thematic Mapper satellite images for 1985, 1993, and 2000/2001. Jeff Wilson in the Department of Geography at IUPUI performed the classification of the LandSat data, producing the datasets classified into 12 land cover categories for the 30-meter pixels.

The procedure for inferring land use was developed using the land cover data for Marion County for 2000 and the land use data for Marion County from the Indianapolis Division of Planning. Measures derived from the land cover data and other datasets are used first to classify land as urban or not urban and then to classify the urban land as residential or employment-related.

Defining Land Uses

The procedure begins with the definition of land uses for the analysis. This includes not only the specification of land use categories within Marion County but the identification of methods to perform such specification for the application of the procedures in the remainder of Central Indiana. Four land use categories are defined: residential land use, employment-related land use, land available for development, and other land uses not available for development.

Residential land use is defined as including all land within the various residential categories in Marion County. Employment-related land use is defined as including the commercial and industrial land use categories plus the hospital, schools, churches, utilities, and other special use categories and the mixed land use category. Urban land use is defined as the sum of residential and employment-related land use.

The Marion County land use data are parcel-based. Streets and roads do not have land uses designated. Since the 30-meter pixel data from LandSat do not allow such features to be consistently resolved, the areas of residential and employment-related land use are extended to the street centerlines (using the ArcView Spatial Analysis “nibble” operation on a grid theme of land use). Because of its large area and unusual land cover, Indianapolis International Airport was excluded from the classification analysis, with the areas of concentrated activity added back as employment-related land use in the final classification.

Land designated on the Marion County land use map as vacant or agricultural are considered to be available for development, with a few exceptions as described below.

The other land uses not available for development and the methods to be employed in defining those areas for both Marion County and the remainder of the region are as follows:

- Interstate and other limited-access highways and their interchanges—for Marion County, appropriate areas separating land uses; for the remainder of the region, the Etak interstate theme plus interchange roads, converted to a 30-meter grid, followed by an "expand 4, shrink 4" operation
- Railroads—for Marion County, railroad land use; for the remainder of the region, the Etak railroad theme, converted to grid, followed by an "expand 1, shrink 1" operation
- Parks and other public lands—for Marion County, park land use; for remainder of region, union of Indiana Department of Natural Resources managed lands theme and Etak parks theme, with water subtracted out
- Water—union of land cover water area and Etak water area and, in Marion County, land use map water area
- Cemeteries—for Marion County, cemetery land use; for remainder of region the removal of large cemeteries by hand in major urban areas after classification
- Large areas of excavation land cover—areas of excavation land cover grouped using "region group" operation having greater than 45 pixels (areas greater than 10 acres)

Performing the Classification

A form of supervised classification was used, employing the information on the residential and employment-related land use and vacant land available for development in Marion County from the Marion County land use map as ground truth (with other land not available for development and the airport excluded). The procedure involved the development of large numbers of classifiers based upon the land cover data and other data followed by the use of stepwise logistic regression to derive the classification. This was done in two steps, first to distinguish urban (residential plus employment-related) land use from nonurban available for development and then to distinguish residential from employment-related land use for those areas classified as urban.

Potential classifiers were developed that were both continuous variables and dichotomous variables. Criteria for selection included not only overall classification accuracy but the reduction of errors in classification to one use or the other. The judgments in the selection of classifiers included the proportional reductions of error assuming higher probabilities of pixels being classified correctly in one direction or the other.

Classification of Urban Land Use

A total of 52 classifiers were developed for the urban nonurban classification. The following continuous classifiers were derived from the land cover data:

- Natural logarithm of number of pixels of high- or low-intensity development in neighborhoods with sizes ranging from 3x3 to 15x15
- Natural logarithm of region size for areas not high- or low-intensity development (using "region group" operation)

The following dichotomous classifiers were derived from the land cover data:

- Numbers of pixels high- or low-intensity development in neighborhoods of various sizes above specified cutoffs, e.g., 10 or more in a 7x7 neighborhood
- Development plus non-development region size less than *n* pixels, e.g., 500
- Areas of high- or low-intensity development applying "expand *n*" operation (*n* = 1, 3, 5, 7, 9)
- Areas of high- or low-intensity development applying "expand *n*, shrink *n*" operation (*n* = 1, 3, 5, 7, 9)

The following continuous classifiers were derived from ancillary data:

- Natural logarithm of population density from census block data
- Natural logarithm of housing unit density from census block data
- Natural logarithm of block size from census block data

The following dichotomous classifiers were derived from the ancillary data:

- Population density over 3.5, 1.5, 0.5, and 0.25 persons per acre
- Housing unit density over 1.5, 0.5, 0.25 units per acre
- Block size 46, 81, 115, 154, 189, 291 pixels or less
- Etak road network, converted to grid, applying "expand *n*, shrink *n*+1" operation (*n* = 5, 6, 7, 8, 9, 10) (The idea for this measure comes from Morisette, *et al.* 1996.)

Stepwise logistic regression in SPSS, with the dependent variable being urban/nonurban in Marion County, was performed to select the classifiers, adding classifiers until the classification accuracy as measured by the percent correctly classified no longer increased. The set of classifiers identified as performing the best classification were:

- Log number developed pixels in 3x3 neighborhood
- Log number developed pixels in 11x11 neighborhood (best classifier)
- Developed pixels expand 9
- Housing density > 0.5 units/acre
- Road network expand 6, shrink 7
- Road network expand 9, shrink 10

The following table shows the classification accuracy:

Reference Marion County	Classified		Total	Percent Correct
	Nonurban	Urban		
Nonurban	142,734	82,572	225,306	63.35%
Urban	56,129	722,716	778,845	92.79%
Total	198,863	805,288	1,004,151	
Percent Correct	71.78%	89.75%		86.19%

For the urban/nonurban classification, 86.2 percent of the pixels were classified correctly. The kappa value for the classification is 0.663. The following table shows the classification accuracy associated with alternative classifications:

Classification	Percent Classified Correctly
High- and low-intensity development land cover	64.8%
1 pixel in 3x3 neighborhood (LUCI 1.0)	81.2%
Ancillary classification (3 classifiers)	83.2%
Land cover classification (1 classifier)	85.1%
Final classification (6 classifiers)	86.2%

Just using the high- and low-intensity development land cover as the classification does a very poor job of distinguishing urban and nonurban pixels. The simple classification, considering a pixel as urban if 1 pixel in the 3x3 neighborhood as urban, which was used for LUCI, does much better. But using the additional information results in distinctly better classification.

This is the classification that was used to classify areas as urban versus nonurban for Central Indiana for 2000. However, the ancillary classifiers are based on year 2000 data and comparable data are not available for the earlier years with land cover data, 1985 and 1993. Therefore, a classification using only the land cover classifiers was also be developed. In this case, a single classifier, the log of the number of high- and low-intensity development pixels in the 11x11 neighborhood provided the best classification, with 85.1 percent of the pixels being classified correctly. This classification was used to estimate urban land for 1985 and 1993, and was also applied to 2000 for use in calculating change, so that a consistent classification was used when calculating the change measures.

Classification of Residential and Employment-Related Land Use

A similar approach is taken to perform the further classification of those areas classified as urban as either residential or employment-related land use. The first attempt employed both land cover and ancillary classifiers. The thought was that the presence of population or housing units might be good indicators of residential land use. And indeed, two measures associated with population density were selected for the classification.

However, when attempts were made to apply the classification in Central Indiana, it was found that this classification procedure resulted in far too many areas being classified as employment-related land use, primarily on the edges of urban areas. The problem arises because at the fringes of urban areas there are large blocks with correspondingly lower population densities. This results in the misclassification of those areas as employment-related land use.

To avoid this problem, the classification of residential versus employment-related land use was performed using only the land cover classifiers. The first set of classifiers use the amounts of land classified as high-intensity development:

- Numbers of pixels high-intensity development in neighborhoods 3x3 to 11x11 (not transformed)
- Whether the numbers of pixels high-intensity in neighborhoods were above certain cutoffs
- High-intensity "expand *n*" (*n* = 1, 2)
- High-intensity "expand *n*, shrink *n*" (*n* = 1, 2, 3)

The second set of classifiers focused on those areas with high-intensity development with a region size (using the "region group" operation) greater than or equal to 15 pixels:

- High-intensity region size, region size >= 15
- Numbers of pixels in neighborhoods 3x3 to 11x11 (not transformed)
- Whether the numbers of pixels in neighborhoods were above certain cutoffs
- "Expand *n*" (*n* = 1, 2, 3)
- "Expand *n*, shrink *n*" (*n* = 1, 2, 3, 5)

A third set of classifiers was based upon the amount of land cover not in high- or low-intensity development:

- Not high-intensity region size <= 45
- Numbers of not high- or low-intensity pixels in neighborhoods 3x3 to 11x11 (not transformed)
- Whether numbers of not high- or low-intensity pixels in neighborhoods were below certain cutoff

Finally, several classifiers used other land cover information:

- Forest and herbaceous land cover
- Low-intensity development, forest, and herbaceous land cover
- High-intensity development and excavation land cover

As before, stepwise logistic regression was used to select the set of classifiers that produced the highest classification accuracy as measured by the percent of pixels classified correctly. The classifiers selected were:

- Number of high-intensity pixels in regions with greater than or equal to 15 high-intensity pixels in 9x9 neighborhood

- Number of pixels not high- or low-intensity development in 7x7 neighborhood
- High-intensity development “expand 3, shrink 3”
- Number of high-intensity pixels in 5x5 neighborhood
- Number of high-intensity pixels in regions with greater than or equal to 15 high-intensity pixels in 11x11 neighborhood
- High-intensity pixels in regions with greater than or equal to 15 high-intensity pixels “expand 3”

The following table shows the classification accuracy for classifying the urban pixels:

Reference Marion County	**Classified**		**Total**	**Percent Correct**
	Residential	**Employment**		
Residential	458,243	33,983	492,226	93.10%
Employment	77,280	153,210	230,490	66.47%
Total	835,523	187,193	722,716	
Percent Correct	85.57%	81.85%		84.60%

For the urban/nonurban classification, 84.6 percent of the pixels were classified correctly. The kappa value for the classification is 0.692.

Model Data

This section describes the major sources of data used in the model and the procedures used to create the data for the Central Indiana Implementation. This implementation uses data for three sets of areal units: The simulation zones are 1-mile-square grid cells. There are 17,369 simulation zones in the region. The employment zones are the 320 ZIP code areas for ZIP codes with spatial extent in the region. The larger subareas are the 44 counties in the region.

Complete definitions of the data files required for the model are provided in the document *luci2 Data Files: Documentation and Instructions for Preparation*.

Land Use Data

The procedures for the classification of urban/nonurban and residential/employment-related land use described above were used for the development of the land use data.

The procedure starts by excluding those pixels considered not urban and not available for development:

- Interstate and other limited-access highways and their interchanges—Etak interstate theme plus interchange roads, converted to 30-meter grid, followed by an “expand 4, shrink 4” operation
- Railroads—Etak railroad theme, converted to 30-meter grid, followed by an “expand 1, shrink 1” operation
- Parks and other public lands—union of Indiana Department of Natural Resources managed lands theme and Etak parks with water subtracted out
- Water—union of land cover water area and Etak water
- Large areas of excavation land cover—areas of excavation land cover grouped using “region group” operation, with more than 45 pixels (areas greater than 10 acres)

The classification procedures were then used to classify the remaining pixels as urban or nonurban. The single land cover classifier was used to perform the classification for the 1985, 1993, and 2000 land cover datasets for determining change. The full classification was applied to classify the 2000 land cover dataset to provide the starting amounts of land for the simulation.

For the pixels classified as urban in 2000 using the full classification, the residential/employment-related land use classification was applied to identify residential and employment-related land use in 2000.

Next, modifications were made to 2000 residential and employment-related land use themes for 2000. For the area around Indianapolis International Airport, the approximate areas including employment-related land use—the terminal, the maintenance areas on the north side, and the FedEx area on the south side—were designated as employment-related land use. The remainder of the airport remains designated as not available for development.

For Grissom Air Force Base, the areas of the runways had been classified as employment-related land use. These areas were changed to the not available for development category.

Large-scale street maps were obtained for each of the major urban areas. Using the Etak roads theme as the base, head-up digitizing was used to exclude otherwise nonurban land as not available for development areas identified on these maps as parks, cemeteries, golf courses, other special uses, and airports. These areas were digitized to form a theme of excluded areas which were then intersected with the urban theme to exclude those areas not urban that are not available for urban development.

Amounts of urban change in the two periods are determined by identifying the pixels that were converted from nonurban to urban use over the period. (The small number of pixels changing from urban to nonurban is ignored.)

Amounts of residential change in the two periods are determined by first subtracting those pixels classified as employment-related in 2000 from the land classified as urban at each of the three points in time (using the simple classification of urban for 2000). Then residential change is the conversion of pixels from nonurban to residential use over each period. Likewise, the amount of employment-related land use change from 1993 to 2000 was estimated by assuming that pixels that were classified as nonurban in 1993 and employment-related in 2000 represented conversion of land to employment-related land use.

To determine the amounts of land available for development in the earlier years, the land not available for development in 2000 (including that land specified by the manual adjustments) was subtracted from the land classified as nonurban in 1985 and 1993.

Finally, the original definition of urban land as consisting of residential plus employment-related land use leads to some inconsistencies in referring to urban land in the model. So urbanized area and urban center/cluster boundary files were used to distinguish land that is not available for development within urban areas from land that is not available outside of urban areas. The land not available within urban areas is now denoted as a third category of other urban land use. Note that this change does not affect changes in urban land use, as land not available for development is not considered to change over time.

All of this information, which is at the pixel level, was aggregated to the mile-square grid cells to get the amounts/proportions of land available for development, urban, residential, employment-related, and not available for development and the measures of change.

Employment Data

Data on employment by ZIP code for 1995 and 2000 have been developed for the 320 ZIP code areas in the region. For ZIP codes partially-inside and partially-outside the 44-county area, the decision whether to include the ZIP code was based upon a judgment as to whether the location of the majority of the ZIP code employment was within or outside the area.

The creation of the employment by ZIP dataset begins with special tabulations of employment by major industry group by ZIP code for 1995 and 2000 that were produced by the Indiana Business Research Center from the ES-202 workforce data compiled by the Indiana State Department of Workforce Development. The industry groups used represent a compromise between preserving detail and minimizing disclosure problems. These industry groups were:

> Rural industries-agricultural, forestry & fisheries and mining
> Construction
> Manufacturing
> Transportation, communications & public utilities and wholesale trade
> Retail trade
> Finance, insurance & real estate
> Services
> Public administration

Less than 5 percent of the industry employment detail was missing due to disclosure.

Data from *ZIP Business Patterns* (*ZBP*) were used to estimate these missing values (U.S. Bureau of the Census 1998, 2003). The midpoints of the employment categories for the establishment by employment data were used to estimate employment by industry by ZIP codes. (For 2000, standard major category NAICS to SIC correspondence was used to convert the *ZBP* data to the major SIC industry groups being used.) These *ZBP* estimates were then used to estimate the nondisclosed industry employment.

For cases in which total employment and public administration employment were known, the *ZBP* estimates were entered for the non-disclosed industries and proportionately adjusted so that total employment across the industry groups matched total ZIP employment. When public administration were not disclosed, the *ZBP* estimates were entered for the other non-disclosed industries and a value for public administration employment was determined so that the total of industry employment matched total ZIP employment. For the relatively small number of cases in which total ZIP employment was not disclosed, the total ZIP employment was used if available or the total ZIP employment estimated from establishment by employment size data was used.

Employment for ZIP codes without spatial extent was added to the spatially-extensive ZIP codes in which the point locations of those ZIP codes were located. This produces the dataset of employment by major industry group by 320 ZIP codes for 1995 and 2000.

Other Model Data

This section describes other data used in the model.

Population Data

Populations by grid cell are needed for the estimation of the density model and for aggregation to the ZIP code areas for the employment change models. Census block data for 1990 and 2000 were converted (as densities) to the pixel level and aggregated to the grid cells.

Transportation Infrastructure

The locations of interstate highways, four-lane highways, and other transportation facilities were extracted from the ETAK highway themes for counties. Distances from the grid cells and ZIP codes to the transportation facilities were calculated.

Distances to routes that might be associated with the construction of new interstate highways and light rail facilities were also calculated for use in the model scenarios.

Utilities Infrastructure

The grid cells served with public water and sewer services were determined for this project. For Marion County, a digital file with the areas provided with sewer service was used. For the remainder of the area and for water utilities, contact was made with the various water and sewer providers to determine the areas served. This information was entered on a grid-cell basis, with a grid cell being indicated as served if any portion of the grid cell was served.

School Quality

Information relating to school quality by school district was obtained from the Indiana Department of Education website, including ISTEP total battery scores 1999-2000, total expense per pupil, 3-yr average 1999, and SAT scores composite 1998-1999. These were converted to the grid cells based on the school district serving the center of each grid cell.

Special Land Data

This section describes data used in the model for scenarios involving special land.

Wetlands

The extent of federally-identified wetlands areas was obtained from the National Wetlands Inventory.

Riparian Buffers

Riparian buffers were created around streams and bodies of water that might be imposed for environmental and flood protection. Buffers were created as follows: 400 foot buffers around the Wabash and White Rivers, the two major rivers in the region; 250 foot buffers around the other major streams and bodies of water appearing in the 1:2,500,000 scale ArcUSA theme from ESRI; and 150 foot buffers around the remaining streams and bodies of water in the 1:100,000 theme from USGS/TIGER files.

Agricultural Land

Agricultural land was determined from the data on land in agricultural use as identified from LandSat data for 1985, 1993, and 2000, with land being considered agricultural if the land was classified in agricultural use at any one time period.

Steep Slopes

Land with slopes greater than 15 percent was determined using a USGS digital elevation model.

Forests

Land in forests in 2000 was obtained from the land cover data. The "region group" command was used to identify areas of forest greater than 20 acres, which are included in the sensitive lands scenario.

Intersections

Because restrictions on land available for development may be applied to more than one category of land, intersections among categories were determined to avoid double-counting in restricting land. Intersections among each pair of land types were obtained for use in the model. The assumption is made that higher-level intersections among three or more land types are sufficiently small and can be ignored.

Model Estimation

This section describes the various models that have been developed and estimated for the prediction of residential development, employment change and employment-related development, and the journey to work.

Probability of Residential Development Model

This model predicts the probability that vacant, available land in grid cells will be converted to residential land use during a simulation period.

Model

An aggregate logit model is used to predict the probability of available land being converted to residential use. The dependent variable is the logit of the proportion of the land actually converted,

$$\operatorname{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right)$$

where p_i is the proportion of land converted to residential use. The equation for the aggregated logit model is then

$$\operatorname{logit}(p_i) = \beta_0 + \sum_k \beta_k X_{ik}$$

where X_{ik} is a set of k predictors of urbanization for each of the i grid cells. Such a model should be estimated using weighted least squares (Wrigley 1985: 32).

Following Wrigley (1985, p. 31) citing Fox (1970), with significant numbers of cases in which the probability is zero or one, a modified form of the logit can be used to avoid losing these cases. This revised form for the logit is

$$\operatorname{logit}(p_i) = \log\left(\frac{r_i + 0.5}{n_i - r_i + 0.5}\right)$$

where r_i is the number of pixels in the grid cell converted to residential use and n_i is the total number of pixels of land available for conversion to residential use in the grid cell. For those cases in which the land available for development is zero, no possibility of residential development exists and the case is treated as missing.

Again following Wrigley (1985, p. 32), such an aggregated logistic model should be estimated using weighted least squares (WLS), with the weight being as follows:

$$w_i = n_i p_i (1 - p_i)$$

This weight will have a value of zero when the probability is *0* or *1*. To avoid losing these cases, Wrigley cites Berkson (1953, 1955) to suggest substituting $0.5/n_i$ instead of *0* when $p_i = 0$ and $1 - 0.5/n_i$ instead of *1* when $p_i = 1$. Then, for those cases in which $p_i = 0$ or $p_i = 1$, the weight will be:

$$w_0 = w_1 = n_i \left(\frac{0.5}{n_i} \right) \left(1 - \frac{0.5}{n_i} \right) = 0.5 \left(1 - \frac{0.5}{n_i} \right)$$

Variables

This section describes the variables included in the model to predict the probability of available land being converted to residential use.

Logit proportion land converted—LGRES90C. This is the dependent variable. This starts with the increase in the amount of residential land from 1993 to 2000, estimated as described above. Because this is the amount of land converted over an eight-year period, this was reduced by 0.625 to provide an estimate of the amount of land converted over a 5-year period, consistent with the length of the simulation period in the model. To determine the proportion of land converted, this was divided by the amount of available land at the start of the period, in 1993, reduced by the amount of land converted to employment-related land use during the period 1993 to 2000. The basis for this reduction was the assumption that new employment-related land development would pay higher prices for available land, reducing the amounts of land available for residential development. This is consistent with the manner in which development will be simulated in the model, with employment-related development being determined first in each simulation period, reducing the amount of available land for the subsequent simulation of residential development. As explained above, the logit of this proportion is the dependent variable.

Log accessibility to employment in 1995—LXEMP99. Accessibility to employment by ZIP code in 1995 was calculated using the following formula:

$$A_i = \sum_j E_j e^{-\beta d_{ij}}$$

where E_j is employment in ZIP code j, d_{ij} is the distance from the center of grid cell i to the point location of ZIP code j, and β is an empirically-determined accessibility coefficient. For this model, of value of 0.000099 for the accessibility coefficient produced the best fit of the model. The natural logarithm of accessibility is used in the model.

Accessibility to employment change—XFMC23. This is accessibility to employment change from 1995 to 2000 as predicted by the employment change models, for consistency with the simulation model. It is calculated in a manner analogous to accessibility to all employment. The accessibility coefficient used for this measure is 0.00023, which produced the best fit of the model.

Sewer service dummy—DSEWER. This is a dummy variable specifying whether the grid cell was provided with sewer utility service.

Water service dummy—DWATER. This is a dummy variable specifying whether the grid cell was provided with water utility service.

Dummy variables for distance to nearest interstate interchange—INCH01C, INCH023. These are dummy variables associated with the distances to the nearest interchange. The first indicates a distance less than or equal to 1 mile from the center of the grid cell to the interchange. The second is greater than 1 mile and less than or equal to 3 miles.

Dummy variables for distance to nearest four-lane highway—FRLN01C, FRLN23. These are dummy variables associated with the distances to the nearest four-lane highway. The first indicates a distance less than or equal to 1 mile from the center of the grid cell to the highway. The second is greater than 1 mile and less than or equal to 3 miles.

Log proportion land residential in 3x3 neighborhood and its square—LRES3X3 and LRES33SQ. These are the natural logarithm of the proportion of land residential in the 3x3 neighborhood of grid cells encompassing each grid cell and that value squared. The inclusion of these variables in the model captures the logistic pattern of residential development over time, in which the rate of development starts out slowly in areas with little development, speeds up as the areas becomes more developed, and then tapers off as the areas becomes more fully-developed.

Logit proportion land converted to residential 1985 to 1993—LGRES89C. This is equivalent to the dependent variable, calculated using the amount of land converted to residential use during the preceding period from 1985 to 1993, adjusted to a five-year rate. The inclusion of this variable captures the persistence in the pattern of residential development over time.

ISTEP score—ISTEP. This is the Indiana Statewide Test of Educational Progress (ISTEP) total battery score for 1999-2000 for the school district in which the grid cell is located.

Model Estimation Results

The model was estimated using weighted-least squares as described above. These are the SPSS regression results for the model:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.631[a]	.398	.398	4.36054974

a. Predictors: (Constant), ISTEP, FRLN01C, INCH23C, INCH01C, DWATER, FRLN23C, LGRES89C, LRES33SQ, DSEWER, LXEMP99, XFMC22, LRES3X3

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	214319.4	12	17859.946	939.286	.000[a]
	Residual	323986.3	17039	19.014		
	Total	538305.6	17051			

a. Predictors: (Constant), ISTEP, FRLN01C, INCH23C, INCH01C, DWATER, FRLN23C, LGRES89C, LRES33SQ, DSEWER, LXEMP99, XFMC22, LRES3X3

b. Dependent Variable: LGRES90C

c. Weighted Least Squares Regression - Weighted by WTRES90C

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-8.950	.239		-37.471	.000
	LXEMP99	.352	.018	.209	19.362	.000
	XFMC22	.000	.000	.074	6.663	.000
	DSEWER	.094	.031	.025	3.041	.002
	DWATER	.280	.031	.072	9.053	.000
	INCH01C	.096	.045	.014	2.163	.031
	INCH23C	.044	.028	.011	1.544	.123
	FRLN01C	.159	.028	.037	5.621	.000
	FRLN23C	.152	.027	.039	5.701	.000
	LRES3X3	8.251	.336	.592	24.532	.000
	LRES33SQ	-10.917	.622	-.372	-17.545	.000
	LGRES89C	.056	.006	.076	9.187	.000
	ISTEP	.028	.003	.070	8.524	.000

a. Dependent Variable: LGRES90C

b. Weighted Least Squares Regression - Weighted by WTRES90C

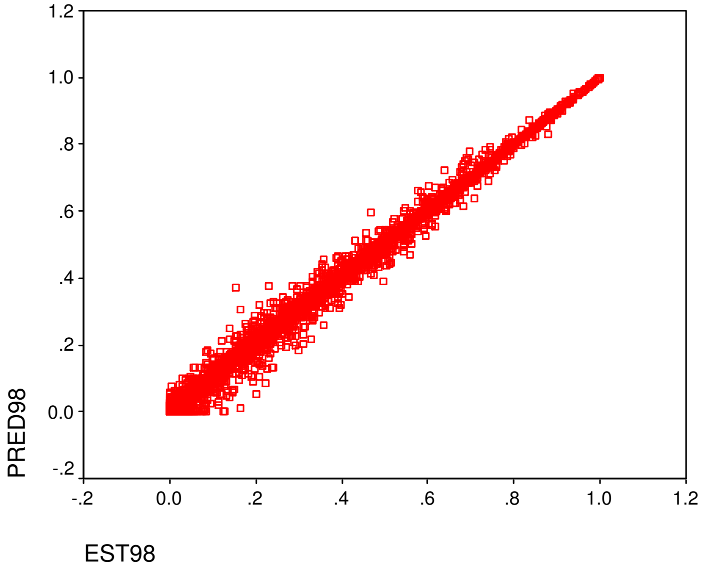
The goodness-of-fit of the model as measured by R-squared is 0.398. All of the predictors are statistically-significant at the 0.001 level.

Model Predictions

The effectiveness of the model in predicting the (known) amounts of residential development from 1993 to 2000 is assessed. The predicted amount of residential development was predicted using this model and the procedures that are used in the simulation.

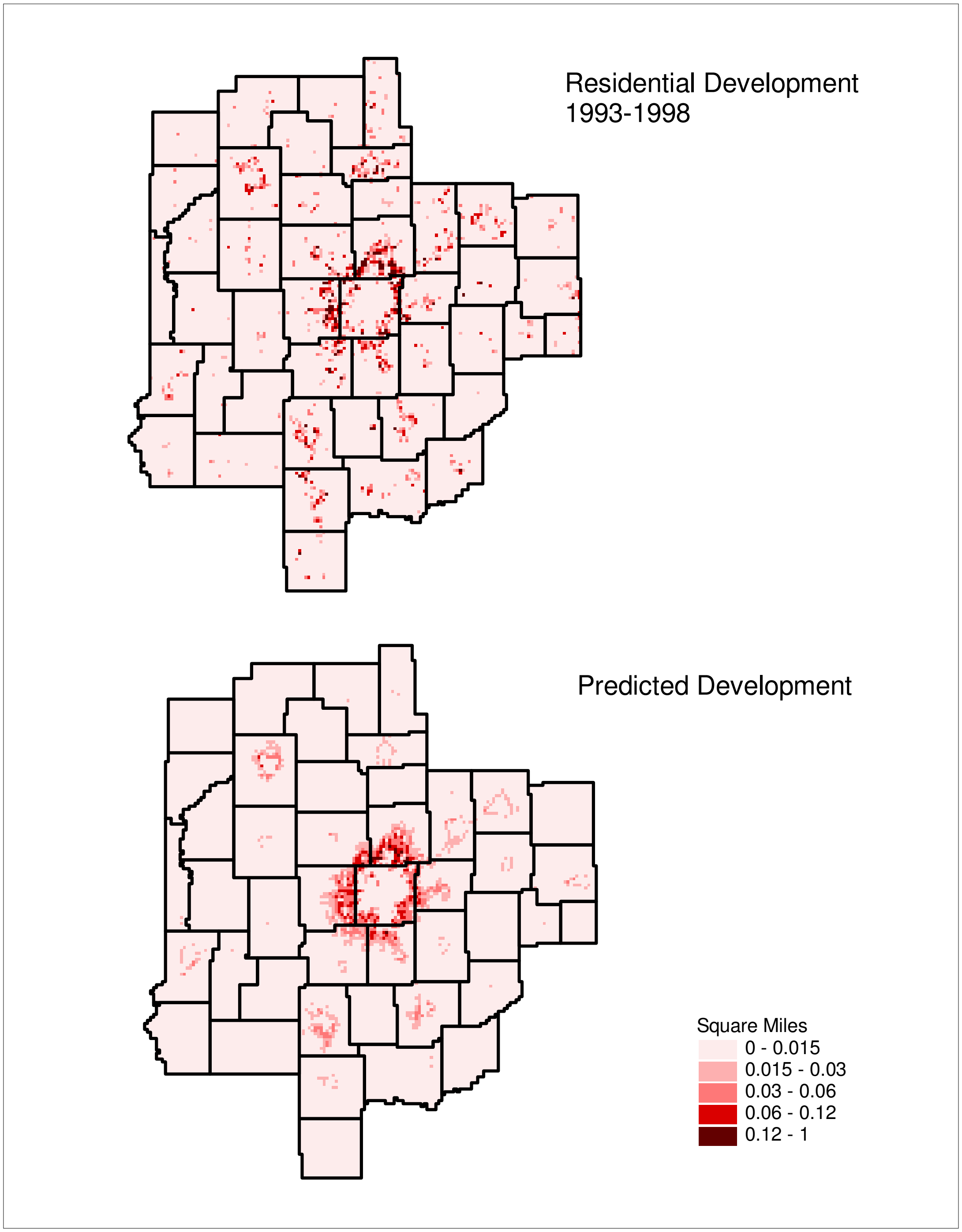
First, the estimated equation was used to predict the logit of the probability of residential development. The values of the predicted logit were adjusted so that the mean and standard deviation of the distribution across the grid cells was the same as for the known distribution. (This adjustment was found to be effective and was used in the *LUCI 1.0* model.) The logits were then converted to the corresponding probabilities. Probabilities less than 0.000277 were set to zero. This is done because the logit model cannot predict zero probabilities. Extremely low probabilities are unlikely to be associated with actual residential development. The value of 0.000277 was selected as this corresponds to the one-half the lowest probability for the grid cells with one pixel of development. The selection was also based on the reasonableness of the overall predictions. This probability was then multiplied by the amount of available land at the beginning of the period to get the initial prediction of residential development. Because these values could result in a total amount of predicted development for the region that would be lower or higher than the actual total development, the probabilities were adjusted up or down using a power transformation in an iterative manner until the predicted total matched the actual total. (In the simulation, predicted probabilities times predicted densities give predictions of the total population that would be accommodated. The probabilities are then adjusted up or down in the same manner to produce an amount of residential development that will accommodate the exogenously-predicted population demand for new residential development in the region.) The predictions are for a five-year period and for the (estimated) amount of residential land in 1998 based upon the 5-year growth amounts used in estimating the model.

The correlation between estimated 1998 residential land use and predicted 1998 residential land use is 0.998. The following plot shows the relationship between predicted and estimated residential land for 1998:



The maps show the estimated and predicted changes in residential land use from 1993 to 1998. The patterns are quite similar, with the predicted pattern showing fewer extremes and a somewhat more evenly distributed pattern of development. This is because the model cannot capture the highly-random variation in development in small areas that occurs in a limited period of time.

Finally, the actual (estimated) and predicted residential development in the grid cells was aggregated to the county level for comparison. *CHRES90C* is the actual development and *DEV1* is the predicted development. Most of the differences are quite small, with the greatest differences occurring in those counties having the greatest amounts of residential development.


Residential Development
1993-1998
Predicted Development
Square Miles
0 - 0.015
0.015 - 0.03
0.03 - 0.06
0.06 - 0.12
0.12 - 1

Report

Sum

CTYNAME	CHRES90C	DEV1
BARTHOLO	1.74873	2.22033
BENTON	.13442	.04407
BOONE	3.07437	2.85328
BROWN	.34426	.53678
CARROLL	.17327	.23789
CASS	.18215	.39907
CLAY	.17670	.35132
CLINTON	.66015	.28217
DECATUR	.08223	.45340
DELAWARE	2.03176	1.94435
FAYETTE	.30142	.31385
FOUNTAIN	.17634	.10530
GREENE	.28976	.45693
HAMILTON	11.06454	10.02254
HANCOCK	2.28573	3.69839
HENDRICK	6.87383	7.28874
HENRY	.55783	.94539
HOWARD	3.32722	1.39500
JACKSON	1.41590	.66457
JENNINGS	.68783	.38497
JOHNSON	2.55260	4.08853
LAWRENCE	2.46090	1.26832
MADISON	2.63395	2.48813
MARION	7.78753	7.83943
MIAMI	1.13893	.55396
MONROE	2.57935	3.20572
MONTGOME	.81891	.62741
MORGAN	3.04626	2.95426
ORANGE	.71688	.19779
OWEN	.09630	.05518
PARKE	.14657	.15606
PUTNAM	.27847	.60508
RANDOLPH	.45226	.13182
RUSH	.26743	.18900
SHELBY	.52835	1.34985
SULLIVAN	.19069	.17761
TIPPECAN	2.66752	2.93746
TIPTON	.30343	.52004
UNION	.64029	.07063
VERMILLI	.14286	.26778
VIGO	.93781	1.75250
WARREN	.14201	.04561
WAYNE	.90175	1.14310
WHITE	.45065	.26343
Total	67.47014	67.48704

Density of Residential Development Model

The second model used in the simulation of residential development predicts the population density of new residential development, establishing the population that will be accommodated by any amount of conversion of land to residential use.

Model

This is a multiple regression model with the log of population density as the dependent variable. Because of the variation in the amounts of residential development across the grid cells, the regression is weighted by the amount of residential land in the cells.

Variables

Log population density—LPOPDEN. This is the dependent variable. It is the natural logarithm of population density by grid cell in 2000, with population from Census 2000, estimated for the grid cells from the block data, and area of residential land in 2000, estimated as described above. This has missing values for cells with no residential land.

Accessibility to employment—LYEMP30. Natural logarithm of accessibility to employment by ZIP code using 2000 employment. The measure of accessibility is as described above in the section on the probability of residential development model. The accessibility coefficient is 0.00030.

Sewer service dummy—DSEWER. This is a dummy variable specifying whether the grid cell was provided with sewer utility service.

Log distance to nearest interstate interchange—LDSTICH2. This is the natural logarithm of distance (in meters) from the center of grid cell to the nearest interstate highway interchange, with 1609 (1 mile) added to the distance before taking the log.

Log ISTEP score—LISTEP. This is the natural logarithm of the Indiana Statewide Test of Educational Progress (ISTEP) total battery score for 1999-2000 for the school district in which the grid cell is located.

Residential land—RES00. This is the variable used for weighting in the regression. It is the amount of residential land in the grid cell in 2000.

Model Estimation Results

The model was estimated using weighted-least squares as described above. These are the SPSS regression results for the model:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.646[a]	.417	.417	.29232463

a. Predictors: (Constant), LISTEP, DSEWER, LDSTICH2, LYEMP30

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	387.620	4	96.905	1134.006	.000[a]
	Residual	541.093	6332	.085		
	Total	928.713	6336			

a. Predictors: (Constant), LISTEP, DSEWER, LDSTICH2, LYEMP30

b. Dependent Variable: LPOPDEN

c. Weighted Least Squares Regression - Weighted by RES00

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	11.517	.456		25.265	.000
	LYEMP30	.209	.007	.384	29.556	.000
	LDSTICH2	-.042	.010	-.046	-4.091	.000
	DSEWER	.591	.022	.312	27.094	.000
	LISTEP	-1.436	.109	-.127	-13.158	.000

a. Dependent Variable: LPOPDEN

b. Weighted Least Squares Regression - Weighted by RES00

The goodness-of-fit of the model, as measured by R-squared, is 0.417. All of the predictors are significant at the 0.001 level.

Employment Change Models

These models predict the changes of employment in ZIP code areas for four major industry groups.

Models

A set of models has been developed to predict employment change by industry across ZIP codes. Models were estimated using the change in employment by ZIP code from 1995 to 2000 as the dependent variable. Predictors considered and included are the following: Accessibility to population and population change as measures of both demand and accessibility to the labor force; distances to interstate highway interchanges; levels of utility service; population, population change, and employment within the ZIP codes; measures of land use within the ZIP codes; and other fixed characteristics within the ZIP codes.

In developing the final set of models, employment in the industry groups retail, service, construction, and public administration were combined into a single local-service employment industry group. These four industry groups showed similar patterns of change and predictors of change. They are logically considered to involve activity primarily serving the local population in the area.

While the models were being developed, it was observed that the dependent variables had extreme outliers that were unduly influencing the regression results. To eliminate this problem, cases in which the value of the industry employment change variable was more than 3 standard deviations above or below the mean were excluded as outliers. Transportation change had one especially extreme outlier that greatly influenced the standard deviation of the entire distribution. In this case, that extreme outlier was dropped first, and then outliers falling more than 3 standard deviations above or below the mean for the remaining cases were excluded as outliers as well.

Local-Service Employment Model Variables

These are the variables in the local-service employment change model:

Change in local-service employment—LOCALCHM. Change in employment in local-service industries—retail, service, construction, and public administration—from 1995 to 2000. This is the dependent variable.

Accessibility to population change—XAPC28E5. Accessibility to population change by ZIP code from 1990 to 1995, estimated as one-half the change in population from 1990 to 2000. This is calculated using the same measure of accessibility described above. The accessibility change coefficient is 0.00028.

Population—POP. Population in ZIP code in 1995, estimated as the mean of the 1990 and 2000 populations aggregated to ZIP code from grid cell (as will be done in the simulation).

Change in urban land—CHURB89. Change in the amount of urban land in the ZIP code from 1985 to 1993 (prior period), aggregated from the grid cells (as will be done in the simulation).

Local-serving Employment Model Estimation Results

The SPSS results for the prediction of change in local-service employment (retail, service, construction, and public administration) are as follows:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.668[a]	.446	.441	483.204601

a. Predictors: (Constant), CHURB89, POP, XAPC28E5

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	58513644	3	19504548.08	83.536	.000[a]
	Residual	72614360	311	233486.687		
	Total	1.31E+08	314			

a. Predictors: (Constant), CHURB89, POP, XAPC28E5

b. Dependent Variable: LOCALCHM

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-44.509	34.923		-1.274	.203
	XAPC28E5	.127	.020	.385	6.323	.000
	POP	.010	.003	.167	3.276	.001
	CHURB89	215.501	64.615	.227	3.335	.001

a. Dependent Variable: LOCALCHM

The goodness-of-fit of the model as measured by R-squared is 0.446. Accessibility to employment change is the most important predictor, which is reasonable. New local-service employment locates in response to new demand. The population of the ZIP code, and the change in the amount of urban land would also seem to be measures related to the demand within the ZIP code.

Manufacturing Employment Model Variables

These are the variables in the manufacturing employment change model:

Change in manufacturing employment—MFGCHM2. Change in employment in manufacturing from 1995 to 2000. The dependent variable.

Square root of population change—SRPOPCH. Change in ZIP code population from 1990 to 1995, estimated as one-half the population change from 1990 to 2000 aggregated to ZIP codes from grid cells (as will be done in the simulation). If population change was negative, this is the negative of the square root of the absolute value of population change.

Manufacturing employment—MFG95. Manufacturing employment in the ZIP code at the start of the period in 1995.

Log distance to nearest interstate interchange—LDSTINCH. Natural logarithm of distance from the ZIP code point to the nearest interstate interchange.

Employment-related land use—EMP93. Employment-related land use at the start of the period, aggregated from the 1993-employment-related land use by grid cell (as will be done in the simulation).

Available land—AVL93. Amount of land available for development at the start of the period, aggregated from the 1993 available land by grid cell (as will be done in the simulation).

Proportion housing units built before 1940 times proportion land urban—ZBEFURB. Proportion of housing units in the ZIP code built before 1940 times the proportion of the land in the ZIP code urban in 1993 aggregated from the grid cells. The idea behind this variable is that it is a measure of areas of old urban development.

Accessibility to population—XAPP29E5. Accessibility to the population by ZIP code in 1995, estimated as the mean of the 1990 and 2000 populations. This uses the same measure of accessibility described above. The accessibility coefficient is 0.00029.

Manufacturing Employment Model Estimation Results

The SPSS results for the prediction of change in manufacturing employment are as follows:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.523[a]	.273	.257	341.543808

a. Predictors: (Constant), XAPP29E5, SRPOPCH, AVL93, MFG95, LDSTINCH, ZBEFURB, EMP93

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13349052	7	1907007.485	16.348	.000[a]
	Residual	35462260	304	116652.173		
	Total	48811313	311			

a. Predictors: (Constant), XAPP29E5, SRPOPCH, AVL93, MFG95, LDSTINCH, ZBEFURB, EMP93

b. Dependent Variable: MFGCHM

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	775.573	209.121		3.709	.000
	SRPOPCH	-4.257	1.053	-.238	-4.044	.000
	MFG95	-.099	.018	-.406	-5.387	.000
	LDSTINCH	-82.264	21.421	-.236	-3.840	.000
	EMP93	282.433	40.953	.684	6.896	.000
	AVL93	1.700	.517	.196	3.285	.001
	ZBEFURB	-.101	.031	-.210	-3.260	.001
	XAPP29E5	-.004	.001	-.263	-2.958	.003

a. Dependent Variable: MFGCHM

The goodness-of-fit of the model, as measured by R-squared, is 0.273. Log distance to interstate is negative and significant, indicating the importance of interstate accessibility. With respect to land use, the amount of employment-related land use and the amount of available land are both significant indicators of the attractiveness of the ZIP code for new manufacturing employment. Change in manufacturing employment is inversely related to the amount of manufacturing employment, indicating the tendencies for employment in large established manufacturing areas to decline. More decline is also expected in older urban areas. New manufacturing tends to locate in areas with lower levels of population change and lower accessibility to population, indicating perhaps a desire for low-cost, uncongested locations.

Transportation, Communications and Wholesale Employment Model Variables

These are the variables in the transportation, communications, and wholesale employment change model:

Change in transportation, communications, and wholesale employment—TRANCHM2. Change in employment in transportation, communications, and wholesale from 1995 to 2000. The dependent variable.

Change in urban land—CHURB89. Change in the amount of urban land in the ZIP code from 1985 to 1993 (prior period), aggregated from the grid cells (as will be done in the simulation).

Log distance to nearest interstate interchange—LDSTINCH. Natural logarithm of distance from the ZIP code point to the nearest interstate interchange.

Employment-related land use—EMP93. Employment-related land use at the start of the period, aggregated from the 1993-employment-related land use by grid cell (as will be done in the simulation).

Transportation employment—MFG95. Transportation employment in the ZIP code at the start of the period in 1995.

Proportion land available—PAVL93. Proportion of the land available for development at the start of the period, aggregated from the 1993 available land by grid cell (as will be done in the simulation).

Accessibility to population change—XAPC22E5. Accessibility to population change by ZIP code from 1990 to 1995, estimated as one-half the change in population from 1990 to 2000. This is calculated using the same measure of accessibility described above. The accessibility change coefficient is 0.00022.

Transportation, Communications and Wholesale Employment Model Estimation Results

The SPSS results for the prediction of change in transportation, communications, and wholesale employment are as follows:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.487[a]	.237	.222	109.227689

a. Predictors: (Constant), XAPC22E5, TRANSP95, LDSTINCH, PAVL93, EMP93, CHURB89

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1116220	6	186036.622	15.593	.000[a]
	Residual	3591137	301	11930.688		
	Total	4707357	307			

a. Predictors: (Constant), XAPC22E5, TRANSP95, LDSTINCH, PAVL93, EMP93, CHURB89

b. Dependent Variable: TRANCHM2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.579	67.427		.083	.934
	CHURB89	-46.447	16.597	-.246	-2.799	.005
	LDSTINCH	-9.259	6.772	-.082	-1.367	.173
	EMP93	81.267	12.376	.574	6.567	.000
	TRANSP95	-.057	.019	-.226	-2.981	.003
	PAVL93	.797	.295	.187	2.701	.007
	XAPC22E5	.021	.004	.378	4.719	.000

a. Dependent Variable: TRANCHM2

The goodness-of-fit of the model as measured by R-squared is 0.237. The amount of employment-related land use is the most important predictor, indicating either the presence of agglomeration economies or location in areas receptive to such uses. Accessibility to population change is significant, indicating this industry shares some characteristics with local-service employment. Change in urban land use is inversely related to employment change, perhaps suggesting an aversion to growing, more expensive and congested areas, while the percent of land available is positively related to change. The amount of employment in transportation, communication, and wholesale is inversely related, suggesting a tendency on the part of larger, more established centers of such employment to grow more slowly. Finally, while it was not quite statistically significant, log distance to the nearest interchange shows the expected negative relationship.

Finance, Insurance and Real Estate Employment Model Variables

These are the variables in the finance, insurance, and real estate employment change model:

Change in finance, insurance, and real estate employment—FIRECHM. Change in employment in finance, insurance, and real estate from 1995 to 2000. The dependent variable.

Population—POP. Population in ZIP code in 1995, estimated as the mean of the 1990 and 2000 populations aggregated to ZIP codes from grid cells (as will be done in the simulation).

Population change—POPCH. Change in ZIP code population from 1990 to 1995, estimated as one-half the population change from 1990 to 2000 aggregated to ZIP codes from grid cells (as will be done in the simulation).

Finance, insurance, and real estate employment—MFG95. Finance, insurance, and real estate employment in the ZIP code at the start of the period in 1995.

Total employment—TOTEMP. Total employment in the ZIP code at the start of the period in 1995.

Employment-related land use—EMP93. Employment-related land use at the start of the period, aggregated from the 1993-employment-related land use by grid cell (as will be done in the simulation).

Proportion employment-related land use—PEMP93. Proportion employment-related land use at the start of the period, aggregated from the 1993-employment-related land use by grid cell (as will be done in the simulation).

Proportion provided with sewer service squared—PSEWSQ. Square of the proportion of the grid cells within the ZIP code in which sewer service is provided, aggregated from the grid cells.

Finance, Insurance and Real Estate Employment Model Estimation Results

The SPSS results for the prediction of change in finance, insurance, and real estate employment are as follows:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.552[a]	.305	.289	95.04552310

a. Predictors: (Constant), PSEWSQ, POPCH, TOTEMP, PEMP93, POP, FIRE95, EMP93

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1220580	7	174368.541	19.302	.000[a]
	Residual	2782365	308	9033.651		
	Total	4002944	315			

a. Predictors: (Constant), PSEWSQ, POPCH, TOTEMP, PEMP93, POP, FIRE95, EMP93

b. Dependent Variable: FIRECHM

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6.541	6.913		-.946	.345
	POP	.002	.001	.215	2.159	.032
	POPCH	.032	.005	.396	6.841	.000
	FIRE95	-.203	.029	-.632	-7.008	.000
	TOTEMP	.007	.002	.369	2.838	.005
	EMP93	-34.700	13.058	-.315	-2.657	.008
	PEMP93	-3.346	.876	-.289	-3.817	.000
	PSEWSQ	.013	.003	.408	5.218	.000

a. Dependent Variable: FIRECHM

The goodness-of-fit of the model as measured by R-squared was 0.305. Growth in finance, insurance, and real estate employment was positively related to ZIP code population and population change, and total employment, all of which could be seen as measures of the demand for the services. Employment change was negatively related with both the amount and proportion of employment-related land use, suggesting that concentrations of such uses were not seen as desirable locations. The amount of employment in this sector was a negative predictor, suggesting that larger established concentrations of such employment might grow more slowly. Finally, availability of sewer utility service was a positive.

Summary of Employment Change Models

This section presents several tables comparing and summarizing the results obtained in estimating the four employment change models. First is the goodness-of-fit of the models:

Industry	**R-squared**
Local service	0.446
Manufacturing	0.273
Transportation, Communications & Wholesale	0.237
Finance, Insurance & Real Estate	0.305

Next is a summary of the predictors included in each of the models. (This table lists the independent variables in the order of the variance accounted for, the magnitudes of the beta coefficients.)

Industry	**Predictors**
Local service	Accessibility to population change (+)
	Change urban land (+)
	Population (+)
Manufacturing	Employment land (+)
	Manufacturing employment (-)
	Accessibility to population (-)
	Square root population change (-)
	Log distance nearest interchange (-)
	Percent units built before 1940 times percent urban land squared (-)
	Available land (+)
Transportation, Communications & Wholesale	Employment land (+)
	Accessibility to population change (+)
	Change urban land (-)
	Transportation employment (-)
	Percent available land (+)
	Log distance nearest interchange (-)
Finance, Insurance & Real Estate	Finance, insurance & real estate employment (-)
	Percent served by sewers squared (+)
	Population change (+)
	Total employment (+)
	Employment land (-)
	Percent employment land (-)
	Population (+)
	Proportion served by sewers (+)

The following table lists all of the predictors in the models, comparing which predictors are used for which industry groups:

Predictor	**Local**	**Mfg**	**Transport**	**FIRE**
Accessibility to population change	+		+	
Accessibility to population		-		
Population	+			+
Population change		-		+
Population density				
Industry employment		-	-	-
Total employment				+
Change urban land	+		-	
Percent urban land				
Available land		+		
Percent available land			+	
Employment land		+	+	-
Percent employment land				-
Distance interchange		-	-	
Percent served by sewers				+
Percent built before 1940 times percent urban		-		

Finally, the ability to predict 2000 employment was assessed by comparing the R-squared values obtained when using 1995 employment by industry plus the predicted change to predict 2000 employment with that obtained by simply using 1995 employment to predict 2000 employment:

Industry	R-squared	
	Predicted Using 1995 Employment Plus Predicted Change	Predicted Using 1995 Employment
Local service	0.982	0.982
Manufacturing	0.914	0.889
Transportation, Communications & Wholesale	0.911	0.909
Trans, Comm & Whole excluding 1 outlier	0.949	0.938
Finance, Insurance & Real Estate	0.922	0.910

Amount of Employment-Related Development Model

For the purpose of allocating the predicted increases in employment by ZIP code as employment-related development to the grid cells, two models are developed. This first model predicts the amount of employment-related land required per employee for each ZIP code. The second model predicts for each grid cell the probability of the conversion of land to employment-related land use.

Model

The model predicts the square miles of employment-related land use per thousand employees in each ZIP code. Because the number of employees varies so widely across the ZIP code, producing very different reliabilities in the estimates of these values, the regression is weighted by the total employment in the ZIP code in 2000.

Variables

Log area per thousand employees—LSQMIEMB. This is natural logarithm of the amount of employment-related land use in the ZIP code in 2000 divided by the total employment in 2000. Outliers greater than 3 standard deviations above and below the mean are excluded. This is the dependent variable.

Accessibility to population—XPOP1D3. Accessibility to population by ZIP code in 2000. This is calculated using the standard approach for accessibility. The accessibility coefficient was 0.00013.

Proportion housing units built before 1940 times proportion land urban—ZBEFURB. Proportion of housing units in the ZIP code built before 1940 times the proportion of the land in the ZIP code in 1993 aggregated from the grid cells. The idea behind this variable is that it is a measure of areas of old urban development.

Model Estimation Results

The model was estimated weighting the regression by total employment as described above. These are the SPSS regression results for the model:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.656[a]	.430	.426	49.06300918

a. Predictors: (Constant), ZBEFURB, XPOP1D3

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	534414.0	2	267207.021	111.004	.000[a]
	Residual	707710.6	294	2407.179		
	Total	1242125	296			

a. Predictors: (Constant), ZBEFURB, XPOP1D3

b. Dependent Variable: LSQMIEMB

c. Weighted Least Squares Regression - Weighted by TOTAL00

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.359	.073		-18.575	.000
	XPOP1D3	-2.61E-06	.000	-.267	-5.117	.000
	ZBEFURB	.000	.000	-.473	-9.068	.000

a. Dependent Variable: LSQMIEMB

b. Weighted Least Squares Regression - Weighted by TOTAL00

The goodness-of-fit of the model as measured by R-squared is 0.430. Higher accessibility to employment is associated with lower amounts of land per employee, which is logical. Likewise, older urban areas have less land per employee.

Probability of Employment-Related Development Model

This model predicts the probability of land in a grid cell being converted to employment-related land use. It will be used in the allocation of employment growth in ZIP codes to the grid cells, with constraints that limit the minimum employment-related land use to be allocated to any grid cell and the maximum number of grid cells within a ZIP code to which such uses may be allocated.

Model

The model is an aggregate logistic model used to predict the probability of vacant available land in a grid cell being converted to employment-related land use. It is estimated using the proportion of available land in 1993 converted to employment-related use in 2000, with the amount being reduced to correspond to the 5-year simulation period. The details of the model, the construction of the variables and the weight, and the use of weighted least squares in the estimation are the same as for the probability of residential development model described above.

Variables

These are the variables included in the model:

Logit proportion land converted—LGEMPCHB. This is the dependent variable. This starts with the increase in the amount of employment-related land from 1993 to 2000, estimated as described above. Because this is the amount of land converted over an eight-year period, this was reduced by 0.625 to provide an estimate of the amount of land converted over a 5-year period, consistent with the length of the simulation period in the model. To determine the proportion of land converted, this was divided by the amount of available land at the start of the period, in 1993. As explained above, the logit of this proportion is the dependent variable.

Log accessibility to population—LXPOP21. This is the accessibility to the population by ZIP code in 1995, calculated in the same manner as the other accessibility coefficients.

Sewer service dummy—DSEWER. This is a dummy variable specifying whether the grid cell was provided with sewer utility service.

Water service dummy—DWATER. This is a dummy variable specifying whether the grid cell was provided with water utility service.

Dummy variables for distance to nearest interstate interchange—INCH01C, INCH023. These are dummy variables associated with the distances to the nearest interchange. The first indicates a distance less than or equal to 1 mile from the center of the grid cell to the interchange. The second is greater than 1 mile and less than or equal to 3 miles.

Dummy variables for distance to nearest four-lane highway—FRLN01C, FRLN23. These are dummy variables associated with the distances to the nearest four-lane highway. The first indicates a distance less than or equal to 1 mile from the center of the grid cell to the highway. The second is greater than 1 mile and less than or equal to 3 miles.

Employment-related land use—EMP93. This is the amount of employment-related land use in the grid cell at the start of the period.

Model Estimation Results

The model was estimated using weighted-least squares as described above. These are the SPSS regression results for the model:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.709[a]	.503	.503	3.37713588

a. Predictors: (Constant), EMP93, INCH23C, FRLN23C, INCH01C, DWATER, FRLN01C, LXPOP21, DSEWER

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	196701.8	8	24587.727	2155.864	.000[a]
	Residual	194535.9	17057	11.405		
	Total	391237.7	17065			

a. Predictors: (Constant), EMP93, INCH23C, FRLN23C, INCH01C, DWATER, FRLN01C, LXPOP21, DSEWER

b. Dependent Variable: LGEMPCHB

c. Weighted Least Squares Regression - Weighted by WTEMCHB

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-12.315	.128		-96.565	.000
	LXPOP21	.784	.015	.398	53.282	.000
	DSEWER	.899	.034	.198	26.303	.000
	DWATER	.317	.034	.068	9.308	.000
	INCH01C	.651	.035	.110	18.789	.000
	INCH23C	.118	.026	.027	4.526	.000
	FRLN01C	.697	.029	.161	24.453	.000
	FRLN23C	.499	.029	.117	17.353	.000
	EMP93	2.073	.109	.112	18.952	.000

a. Dependent Variable: LGEMPCHB

b. Weighted Least Squares Regression - Weighted by WTEMCHB

The goodness-of-fit of the model as measured by R-squared was 0.503. Probability of employment-related development increases with accessibility to population, presence of utilities, proximity to highways, and the amount of employment-related land use.

Journey-to-Work Model

A simple double-constrained gravity model operating at the ZIP code level will be used in the model to estimate the mean journey to work.

Model

The journey-to-work estimation uses the number of employees by ZIP code and the number of employed persons by place of residence by ZIP code (obtained by adjusting population aggregated to the ZIP codes so that the total equals the total for employees). Then the following double-constrained gravity model is used to predict travel flows and estimate the mean journey to work:

$$T_{ij} = k_i l_j O_i D_j \exp\left[-\beta\left(d_{ij} + \delta\right)\right]$$

where T_{ij} are the trips from ZIP code i to ZIP code j, O_i are the origins (employees by place of residence) in i, D_j are the destinations (employees by place of work) in j, and d_{ij} is the distance from i to j. The parameters k_i and l_j are the balancing factors used to make the total predicted origins and destinations equal to the actual totals. The parameter β is the distance-function parameter associated with rate of decline of interaction with distance. The parameter δ is a fixed amount added to each distance to reflect terminal times for trips and within-zone travel. Values for these two parameters need to be estimated.

Variables

The following variables are used in estimating the model:

Employees by place of residence (origins). This is the population for 2000, aggregated to the ZIP codes, and adjusted so that the total across the region is equal to the total number of employees by place of work.

Employees by place of work (destinations). This is the employment by ZIP code for 2000.

Trip length distribution for Central Indiana. This is the distribution of trip lengths in minutes for the 44 counties in Central Indiana from the 2000 census.

Mean trip length by ZIP code. This is the mean trip length from the 2000 census for each ZIP code.

County-to-county journey-to-work trips. This is the number of journey to work trips from each county to each county in Central Indiana, from the Census 2000 county-to-county work flows data.

Model Estimation Results

A small program, the *luci2JourneytoWorkTool*, was developed to calculate predicted work trips using the double-constrained gravity model for any parameters and to compare the predicted results with known data on work travel. The program includes options to specify the accessibility coefficient, the fixed amount to be added to each distance, an amount to be added to each distance based upon the square root of the area of the zone (which was not used), and whether intrazonal distances were to be set to be one-half the distance to the nearest zone. This latter option was selected for the estimation of the model, as it produced clearly superior results to leaving the intrazonal distances at zero.

Multiple measures are calculated to assess performance. The first compares the predicted trip length distribution with the trip length distribution in minutes for the entire area from the census. (Predicted trip lengths are converted to the census units in minutes by taking the ratio of the mean trip length in minutes from the census to the mean predicted trip length). The Chi-squared measure is used to assess the extent to which the distributions are similar.

The next two measures compare the predicted mean trip length for each ZIP code area with the census mean trip lengths for those ZIP codes. (Predicted trip lengths are converted to the census units in minutes by taking the ratio of the mean trip length in minutes from the census to the mean predicted trip length). The squared differences between predicted and census mean trips lengths are summed across the zones and the RMSE is calculated as the measure of the error. The first of these two measures calculates the RMSE directly. The second error weights the sum of squared errors by the numbers of origins and destinations in each zone.

The final two measures compare the predicted county-to-county trips, aggregated from the ZIP codes, with the census county-to-county trips. The squared differences between predicted and actual trips across all pairs of origins and destinations are summed and the RMSE is calculated as the measure of the error. The first of these two measures calculates the RMSE directly. The second error weights the sum of squared error by the numbers of origins and destination in each county from the census.

An iterative procedure was followed to estimate the parameters. Different values for the β and δ parameters were tried and the various measures for assessing the performance of the predictions were compared. The following estimated values were obtained:

β *(distance function parameter)* = 0.000148
δ *(fixed amount added to distance)* = 3430 (meters)

Note that the value for the distance function parameter is consistent with values obtained for accessibility coefficients in earlier models. The value of 3430 meters for the fixed amount to be added to distance is just over two miles, which seems reasonable.

The predicted mean journey to work for the estimated model was 14,956 meters or 9.29 miles. The mean journey-to-work time for the region from the 2000 census was 22.60 minutes. So the predicted mean journey to work would imply a travel speed of 24.7 miles per hour, which seems reasonable.

Rate of Decline of Accessibility Coefficient

When the journey to work model was implemented within the *luci2 Urban Simulation Model*, when using the employment growth options involving prediction of local service employment, the model predicted lower values for the mean journey to work at the end of the simulation than the starting value in 2000. The values were lowest with the option to predict both local-service and basic employment. For the other employment growth options, the journey to work model predicted increases in the mean journey to work, though these increases were quite small. This was inconsistent with the observed increase in the mean journey to work for the Central Indian region from 20.52 minutes in 1990 to 22.60 minutes in 2000 as reported by the Census. This is an increase of 10.1 percent.

Considering the nature of the employment predictions, which will tend to locate employment growth near areas of the rapid population growth, this predicted decline in the mean journey to work could be explained. However, this pattern of greater employment growth in the more rapidly growing suburban areas is also what has actually been occurring in metropolitan areas in the region (and much of the United States). Yet despite the fact that new jobs appear to be located closer to residences, the mean journey to work has been increasing significantly. So the suburban residents are not necessarily choosing to work at nearby places of employment, which is not necessarily surprising, as many of these nearby jobs are in the service industry.

If the mean journey to work is increasing at the same time as new jobs are being located closer to places of residence, then the relative importance of accessibility to employment for the journey to work must be decreasing. This implies that the accessibility coefficient in the journey to work model has been decreasing over time and should be predicted to continue to increase for the model.

This requires estimating the rate of decline in the accessibility coefficient for the journey to work model. Two approaches were used to perform this estimation. The first estimated the accessibility coefficient for 1990. The second estimated the accessibility coefficient for 2010 that would result in the simulation model predicting the observed rate of increase of the mean journey to work from 2000 to 2010.

For the estimation of the accessibility coefficient for 1990, data were available on population by ZIP code and, from the Census, the distribution of trips by trip time for the region and the mean journey to work by county. The earliest employment data available by ZIP code were for 1995. So employment by ZIP for 1990 was estimated using the change in employment from 1995 to 2000 (assuming the same changed occurred during the first half of the decade.

One issue was the criterion to be used in estimating the accessibility coefficient for 1990. Given the limitations in the data available, the estimation for 2000 involved some degree of judgment as the alternative criteria available for assessing the predictions made by the journey to work model were somewhat inconsistent. Since the objective was to determine the accessibility coefficient consistent with the observed increase in the journey to work from 1990 to 2000, the choice was made to use the mean predicted trip length. The accessibility coefficient was determined that produced a predicted mean trip length for 1990 such that the mean trip length predicted by the model for 2000 was 10.124 percent greater than the mean predicted for 1990. (This is the percentage increased determined from the Census.) This produced the following estimate for the accessibility coefficient for 1990:

β (distance function parameter) for 1990 = 0.000197
Decrease in accessibility coefficient from 1990 to 2000 = 0.000049
Percentage decrease in accessibility coefficient from 1990 to 2000 = 24.87%

The second approach used the simulation model to simulate development to 2010 using the employment growth option to predict local service and basic employment (the preferred employment growth option with the lowest journey to work prediction). The accessibility coefficient for the calculation of the final mean journey to work was determined that resulted in a rate of increase in the predicted mean journey to work from 2000 to 2010 that was the same as the observed rate of increase from 1990 to 2000, 10.124 percent. This produced the following estimate for the accessibility coefficient for 2010:

β (distance function parameter) for 2010 = 0.0001114
Decrease in accessibility coefficient from 2000 to 2010 = 0.0000366
Percentage decrease in accessibility coefficient from 2000 to 2010 = 24.73%

The amount of increase in the predicted mean journey to work becomes greater as the size of the accessibility coefficient becomes lower. Therefore, it seems reasonable to use the rate of decline for comparing these results. Using that criterion, the results are virtually identical. Because of the problems with the employment data for 1990 and the issues involving the estimation of the accessibility coefficient, the choice is made to use the estimated accessibility coefficient for 2010 in the model.

The final issue involves the determination of the final accessibility coefficient for use in the model for simulation periods other than 2010, assuming continued decline in the coefficient. As mentioned above, using a constant amount of decline of the accessibility coefficient per period results in significantly greater increases in the predicted mean journey to work per period for later years. Assuming a constant rate of decline in the accessibility coefficient improves the situation somewhat, but this still results in greater increases in the predicted mean journey to work from one period to the next for later periods.

The method selected was to use an inverse power function of time to determine the accessibility coefficient for other periods. The formula being used is:

$$\beta_t = \beta_0 (t+1)^p$$

where β_0 is the estimated accessibility coefficient for 2000, β_t is the accessibility coefficient to be used for predicting the mean journey to work at the end of t simulation periods, t is the number of simulation periods, and p is the power such that the 2010 accessibility coefficient is as estimated above. This function has two desirable properties: First, it results in by far the least variation in the period-to-period predictions of the mean journey to work as compared with other methods considered. Second, it produces somewhat greater increases in the period-to-period predictions of the mean journey to work in the first few periods, with the increases declining somewhat and leveling out for the later periods. This seems to be a reasonable and conservative choice.

Model Simulation

Simulation of future changes in employment, employment-related land use, and residential land use proceeds in five-year increments starting with the values for population, employment, and land use in 2000. This section describes the manner in which the basic simulation proceeds. The implementation of the scenario options within the simulation is described in the following section with the exception of several user specifications that are central to the basic simulation.

Population Growth

Change in the model is driven by an exogenous specification of population growth for the entire region that drives the demand for new employment-related and residential development. The first step in each round of the simulation is the calculation of the increase in population as specified by the user. If the option is selected to use a subarea (county) forecast of population as control totals for some future target year, the increase in population in each subarea is calculated by interpolation or extrapolation using the starting and forecast target year populations. Note that if this option is selected, the subarea changes are used as control totals only for the allocation of new residential development, not for employment change or new employment-related development.

Employment Growth

Total employment in the region is assumed to grow at the same rate as the population. This means that any changes in the overall labor-force participation rate are not considered, and that rate is assumed to be constant. It is assumed that over a long period of time this is not an unreasonable assumption.

The user can make a number of choices about how employment changes are to be predicted for the employment zones (ZIP code areas). The simplest options involve allocating the total increase in employment (determined by the population growth rate) to the ZIP codes using basic shift-share methods. For the specification that the growth in employment be allocated in proportion to ZIP code employment, employment in each ZIP code is simply increased by the population growth rate.

The two additional simple options provide for the allocation of growth in employment in proportion to recent, historical growth in either the subareas (counties) or the ZIP codes. In these cases, for those areas, counties or ZIP codes, that had historical increases in employment, the proportions of the total of those increases are determined and the predicted employment growth is allocated using those proportions. When the allocation is to the counties, the growth in county employment is then allocated to the ZIP codes in proportion to the current ZIP code employment.

Note that in these allocations in proportion to recent historical growth, the allocation is made only to those areas that had positive growth. No allocation is made and there is no change in employment in those areas that experienced recent historical declines. This choice was made for several reasons. Declines in employment observed over a short period, especially very large declines, could well be the result of one-time events, such as a plant closing, that might not be expected to be repeated on a regular basis in the future. For areas that experienced very large declines in the recent period, predictions of future declines based upon that could literally result in the employment in the area approaching or reaching zero over the period of the simulation. Finally, the method that might be used to reasonably forecast such declines in employment is not obvious. Certainly allocations of growth in proportion to both historical growth and decline are not reasonable, as this would result in large declines being predicted with greater increases in overall employment growth. The most reasonable approach might be to assume a constant rate of decline in employment. But this will produce the problems of employment approaching zero in the areas with the greatest recent historical decreases. These problems are especially severe for the smaller, ZIP code areas, which have greater variation in employment over short periods.

Alternatively, the user can specify that the models estimated to predict change in employment by ZIP code be used for the allocation of the total predicted employment growth. The options include using these models to predict change in employment in all four industries and using only the model to predict change in local service employment, with simpler methods being used to allocate growth in basic employment. These latter options are included because it is believed that the models to predict employment change in the other industries may be less reliable.

In using these models, the assumption is made that the share of total employment in the local-service industries remains constant. So the total growth in employment is divided into growth in local-service employment and growth in basic employment, which are allocated separately, and both grow at the same rate at the population.

For the prediction of change in local-service employment, that model is used to predict the change in employment by ZIP code. These predicted changes are then proportionately adjusted so that the total of predicted changes by ZIP code is equal to the total growth in local-service employment predicted for the entire region. Predictions of negative changes are used but are not adjusted.

For the prediction of change in employment in the remaining three industries, those models are used to predict the changes in employment by ZIP codes. Then all of these predicted changes by industry by ZIP code are proportionately adjusted so that the total of predicted changes by ZIP code is equal to the total growth in basic employment predicted for the entire region. Note that this means that the proportions of employment in the three industries will not remain constant.

If only local-service employment is predicted by the model, the user can specify various options for allocating predicted growth in basic employment to the ZIP codes. These options are analogous to the simple options for allocating total employment growth: allocation in proportion to ZIP code basic employment and allocation in proportion to recent growth in either county or ZIP code employment. These are implemented in the same manner as for total employment allocation.

The *luci2 Urban Simulation Model* includes the capability of providing for two additional options for determining employment change using external forecasts that are not included in the Central Indiana Implementation. One option allows the user to select a file including a forecast of total employment by employment zone for target year specified for the simulation. Then employment change for each simulation period is obtained by interpolation between the starting and forecast employment. A second option allows the user to specify that local service employment be predicted by the model with basic employment change being determined by an exogenous forecast of basic employment for the employment zones for some year. Then change in basic employment is obtained by interpolation or extrapolation using the starting and forecast basic employment.

Employment Forecasting Results

This section presents and discusses the employment forecasts produced by the Central Indiana Implementation for the simulation year 2025. This uses default recent population growth, all other default values, and no utility expansions. (The scenario options selected affect development and therefore the predictions of local service and basic employment using the models.) The table on the following page presents the forecast populations summarized by county for each of the forecast options.

To provide some basis for comparison and evaluation, the table also includes a forecast derived from the Woods & Poole county employment forecasts for 2023. Since the total employment growth for the region forecast by Woods & Poole is considerably higher than the totals produced by the model, the Woods & Poole county forecasts were adjusted downward proportionately so that the total adjusted Woods & Poole forecast change equaled the (average) model forecast change.

The Woods & Poole county employment forecasts are just that—forecasts—and are not the actual employment levels for 2025. However, examination of the differences between each of the model forecasts and the Woods & Poole forecast offers one way of comparing and evaluating the model forecasts. Taking each set of model forecasts by county, the differences from the Woods & Poole forecasts were calculated and two measures of those differences, the mean absolute deviation (MAD) and the root mean squared error (RMSE) were determined for each model forecast. The percentage differences by county from the Woods & Poole forecast were likewise calculated and the mean absolute deviation and root mean squared errors for those differences were determined for each model forecast. The results are presented in the table on page 53.

Forecast 2025 Employment Change Using Alternative Methods

County	Pr local & basic	Pr local Bas pr zip	Pr local Bas cty gr	Pr local Bas zip gr	Prop to zip	Prop to cty gr	Prop to zip gr	W&P adj total
Bartholomew	53,060	56,396	49,315	49,683	54,090	44,238	44,059	52,849
Benton	4,111	2,836	2,806	2,779	3,502	3,296	3,183	4,264
Boone	32,102	23,169	26,084	24,459	18,065	20,562	19,570	22,023
Brown	5,481	5,384	5,622	5,465	4,254	4,634	4,382	6,065
Carroll	7,157	6,768	6,855	6,633	6,851	6,608	6,606	8,424
Cass	18,893	21,641	25,745	23,270	22,683	26,309	24,724	19,821
Clay	10,086	7,089	6,729	6,768	6,887	5,257	5,613	11,012
Clinton	14,271	11,834	11,061	10,903	12,218	9,506	9,572	14,028
Decatur	18,428	17,400	19,646	17,611	17,902	21,972	20,464	18,917
Delaware	62,419	65,675	63,344	66,489	65,723	58,708	61,053	59,502
Fayette	12,235	13,541	24,249	18,856	13,376	20,686	18,784	12,977
Fountain	9,977	7,267	10,390	8,763	7,831	9,240	8,156	7,218
Greene	7,755	7,300	6,802	7,008	7,762	5,925	6,926	11,757
Hamilton	114,268	111,156	125,518	115,776	97,825	125,567	116,732	122,315
Hancock	32,496	28,644	30,794	29,355	20,532	22,474	21,333	26,049
Hendricks	71,188	58,326	61,907	59,383	44,733	51,643	49,913	54,633
Henry	18,789	18,795	18,622	18,617	18,467	15,563	16,384	18,603
Howard	51,270	60,004	53,029	54,040	60,662	48,173	51,986	52,386
Jackson	29,694	25,979	28,499	26,001	26,204	27,048	25,791	25,376
Jennings	11,248	9,771	11,075	10,143	9,525	7,421	10,064	12,568
Johnson	73,172	61,807	68,186	63,982	54,906	69,911	64,828	57,297
Lawrence	21,421	21,175	23,700	21,918	19,096	18,598	18,142	19,576
Madison	58,372	61,652	56,606	61,628	58,261	44,497	52,082	48,034
Marion	610,354	664,999	637,581	660,843	728,819	726,804	727,102	657,233
Miami	13,063	12,590	12,999	12,287	12,555	12,885	12,377	11,964
Monroe	63,778	68,813	65,177	71,722	64,727	64,158	68,334	79,831
Montgomery	24,223	21,923	19,626	19,630	22,194	19,875	19,540	19,565
Morgan	26,670	23,955	23,279	23,343	17,940	16,890	16,382	23,540
Orange	7,046	7,921	7,103	8,580	8,524	7,315	8,593	7,891
Owen	5,405	4,648	5,555	5,050	4,607	5,601	5,355	6,991
Parke	4,360	3,565	3,268	3,477	3,736	3,596	3,655	4,757
Putnam	19,188	15,908	16,013	15,552	16,731	17,607	17,102	18,829
Randolph	8,165	8,975	7,988	8,622	9,981	9,244	9,245	9,018
Rush	8,412	7,090	8,485	7,755	8,054	9,137	8,594	8,600
Shelby	25,535	23,319	24,182	23,449	20,360	19,842	20,363	22,084
Sullivan	6,656	5,244	6,041	5,781	5,489	5,486	5,703	8,584
Tippecanoe	93,662	88,130	90,483	88,720	83,726	87,576	86,169	96,862
Tipton	6,710	6,835	7,992	7,289	5,679	6,087	5,783	5,400
Union	2,143	1,777	1,944	1,819	1,898	2,072	1,959	2,293
Vermillion	4,743	3,931	3,812	3,902	4,114	4,354	4,116	5,723
Vigo	60,094	57,331	52,958	54,798	58,667	49,390	52,876	56,668
Warren	3,600	1,709	1,975	1,829	1,953	2,476	2,367	2,379
Wayne	42,103	43,559	44,409	42,590	45,238	41,311	40,593	42,194
White	15,970	12,715	11,106	11,974	13,375	10,210	12,197	12,995
Total	1,789,773	1,788,546	1,788,560	1,788,542	1,789,722	1,789,752	1,788,752	1,789,092

Comparison of Model Employment Forecasts by County with Woods & Poole

Model Forecast	Differences in Forecast Employment		Percentage Differences in Forecast Employmnt	
	MAD	RMSE	MAD	RMSE
Predict local and basic	4,230	8,849	14.2	18.7
Predict local, basic prop to ZIP emp	2,904	4,332	13.7	17.9
Predict local, basic prop to county growth	3,630	5,395	17.8	24.1
Predict local, basic prop to ZIP growth	2,813	3,920	15.2	19.5
Emp growth prop to ZIP emp	5,112	12,276	14.7	17.7
Emp growth prop to county growth	4,947	11,529	17.5	22.4
Emp growth prop to ZIP growth	4,478	11,289	14.9	19.3

First consider the differences in forecast employment levels between the model forecasts and the Woods & Poole forecast. Both the mean absolute deviation and the root mean squared error measures show the three forecasts in which the model predicts local service employment and then uses a simple method to allocate growth in basic employment to be closest to the Woods & Poole forecast. The three simple methods of allocating employment growth were the most different. The prediction of both local-service and basic employment was in the middle.

When one looks at the percentage differences in the county forecasts from the Woods & Poole forecast, a rather different picture emerges, with a less consistent pattern. The three model forecasts that were closest to the Woods & Poole forecast were the first two methods and the simplest method, allocating employment growth in proportion to ZIP code employment. Given that they are either low or in the middle on the measures of both employment differences and percentage differences, the first two methods—the model predicting both local-service and basic employment and the method predicting local-service employment and allocating basic employment growth in proportion to ZIP code employment—are perhaps the most consistent and may be the preferred choices.

Methods allocating either basic employment growth or all employment growth in proportion to past growth, either in the county or the ZIP code, may have some significant problems because they are based upon employment-change in a single five-year period. Two counties show the problems in the two directions: Bartholomew County has very low forecast growth, obviously as a result of low levels of employment growth in the recent period. On the other hand, Fayette County has very high forecast growth because of some significant employment growth during the period.

Employment-Related Development

The simulation of residential development begins with the prediction of the number of square miles per thousand employees for each of the employment zones (ZIP codes) using the model described above. This is multiplied by the forecast change in the number of employees to obtain the forecast employment-related land use to be allocated within each employment zone. If the forecast of employment change in a zone is negative, the amount of land to be allocated is set to zero.

The logit of employment-related development is then predicted for each of the simulation zones (grid cells) using the model described above. This is then converted to the predicted probability.

In the simulation of residential development, the predicted logit is then used to determine the amount of residential development to be allocated to each simulation zone. For employment-related development, this is not appropriate as employment land use tends to be more clustered in limited areas. To accomplish this, two parameters are used: The maximum percentage of the simulation zones within an employment zone to which new employment-related development can be allocated and the minimum quantity of new employment-related development that can be allocated to any simulation zone. These are user-adjustable in the Advanced Options section.

The allocation proceeds sequentially for each of the employment zones. The maximum number of simulation zones to which employment-related development can be allocated is determined. Equal amounts of employment-related development will be allocated to each of those zones. That allocation amount is determined. If that amount is less than the minimum quantity of new employment-related development that can be accommodated, the number of zones to which development can be allocated is reduced accordingly.

The set of simulation zones to which the allocation is to be made are those simulation zones having available land that have the highest predicted probabilities for employment-related development. The employment-related development is then allocated to those zones.

A variety of situations can arise that require modifications to this procedure. If the amount of available land in a selected simulation zone is less than the amount to be allocated, then the available land is allocated as employment-related development and the remainder must be reallocated. If possible, this unallocated development is allocated to other simulation zones within the set of simulation zones selected for allocation.

It is possible that all of the available land in the selected set of simulation zones can be allocated for employment-related development without having allocated the predicted amount of development for the employment zone. In that case, the simulation zone with available land with the next highest probability of development is selected and the allocation is made. If land remains to be allocated, the process is repeated until all of the land has been allocated or until all available land in the employment zone has been allocated to employment-related development.

For some of the employment zones, not all of the predicted employment-related development will be allocated. Indeed, in the Central Indiana Implementation, there are 12 ZIP code areas that are fully urbanized with no available land at the start of the simulation. In these cases, the assumption is made that the employment growth will be accommodated by more intensive use of the existing employment-related land. The example of such development in an essentially fully-developed central business district is an example.

After the first round of the simulation, it is possible and likely that some of the high-probability simulation zones to which employment-related development has been previously allocated will have no more available land because of residential development having occurred in that zone. This is made more likely because factors that make a zone attractive for employment-related development also make the zone attractive for residential development. This creates two problems: First, if the entire employment zone runs out of land because of residential development, then less employment-related development will be allocated than otherwise might be expected. Second, even if this does not occur, employment-related development may become more dispersed within the employment zone than desired, as the prime simulation zones for employment-related development run out of land because of residential development.

This is addressed in the allocation by temporarily subtracting prior residential development in each simulation prior to the employment-related allocation, increasing to amount of available land to the initial amount less the employment-related development that has already been allocated in the prior simulation rounds. The employment-related development is then allocated as described above. Then, for each simulation zone, if the amount of employment-related development allocated is greater than the amount of available land including the prior residential development, the additional amount is deducted from the amount of residential development and the population associated with that development is determined and summed for the entire region. That population reduction is then added to the amount of population to be allocated in the next simulation, so the residential development that was deducted is then simulated and added back.

The rationale for this approach is that available land with good prospects for future employment-related development is likely to have higher prices and to be withheld for that purpose rather than being made available for residential development. Obviously, the longer the simulation proceeds, the more tenuous this becomes. However, there is no basis for making the decision to restrict the reversal of residential development to some shorter period. The judgment was made that the procedure being used here represents the most reasonable approach.

Furthermore, the population and residential development affected by this procedure are quite small. In running a simulation to 2025, using default scenario setting with population growth of 5.5 percent per simulation period, population in the region increased by about one million, an average of about 200,00 per simulation period. In simulation periods 2 through 5, the total population reduction resulting from the employment-related development that had to be included in the subsequent residential development simulation ranged from about 3,700 to 4,600 or around 2 percent of the total population being allocated in each period.

Residential Development

The amount of residential development simulated in each population amount is determined by the total demand for new residential development in the region. This is the sum of the forecast population growth for the region and the internal demand for new residential development. The amount of new residential development during any period depends upon both population growth and the number of current residents in the region seeking new development, either as a result of new household formation (and smaller household sizes) or movement from areas of existing residential development. For the Central Indiana Implementation, the level of internal demand per five-year simulation period was estimated using data on new residential development and population growth from 1990 to 2000. It is specified as a percentage of the current population demanding new development. The value of this percentage can be changed by the user.

If the option to use county control totals is selected, no internal demand for new residential development is included.

The simulation of residential development begins with the prediction of the logit of the probability of change in land to residential use using the model estimated above. The distribution of the predicted logit is adjusted to be equal to the distribution of the original distribution for the known change, as described above. The logit is converted to its probability. Probabilities less than the minimum probability cutoff threshold specified in the model are set to zero. This minimum probability threshold can be changed by the user.

The next step is the prediction of the log of population density. This is then converted to the prediction of the population density.

The population that would be accommodated by multiplying the predicted probability of residential development times the land available for residential development times the predicted density is determined and summed for the region (except when using the county subarea control totals). This is compared with the forecast population growth for the region. The probabilities are then adjusted up or down using a power transform until the predicted total population that would be allocated equals the forecast population growth.

The situation can occur in which the predicted population growth remains less than the forecast population growth even after all of the nonzero probabilities are adjusted to one, implying total development of all available land. This means that the simulation has run out of land to accommodate the forecast population growth for the region, at least during this round of the simulation (and this is reported to the user).

In the case in which the simulation has run out of land, the assumption is that the additional population will be accommodated by development at higher densities than predicted or otherwise specified by scenario options. The assumption is that this will have been taking place throughout the simulation in all areas experiencing growth, as developers and residents anticipate the limitations on land available for development and respond to the consequential increases in land prices.

Therefore, when the model reaches the point in which it has run out of land for additional residential development, the additional population that has not been accommodated by development of all available land at the specified densities is then allocated to all simulation zones (grid cells) in which development has occurred in the simulation thus far, in proportion to the share of total population growth in each simulation zone.

When the users has specified the use of subarea (county) forecasts as control totals, the adjustment of the probabilities to accommodate the forecast growth is done separately for each subarea. For each subarea, the possibility of running out of land exists (independent of the other subareas). In such cases, additional population is allocated to simulation zones within the county in the same manner as followed for the region as a whole.

Simulation Round Updates

At the end of each simulation round, certain values have to be updated. The manner in which these updates take place is controlled by the various scenario option choices. Most of these end-of-round updates are discussed in the description of the implementation of the scenario options. However, the updating of the areas provided with utility service has the potential to affect most simulations and is discussed here.

Depending upon the utilities scenario option choices, water and sewer utilities can be expanded at a specified rate and/or automatically as the simulation zones receive more urban development. The updating and expansion of the utility service areas is described here.

The user can specify that utilities be expanded around the major urban areas so that area served at the end is approximately some specified percentage of the current urban areas at the start of the simulation. This is implemented by having a file for the simulation zones with each simulation zone having the cumulative percentage of the areas surrounding and closer to the urban areas. The value will be 0 for zones within the major urban areas and x percent ($x > 100$ percent) for that simulation zone for which the area of that zone plus all other zones closer to the urban area plus the urban area is percent of the urban area.

It is assumed that such growth occurs as the uniform expansion beyond the boundary of the existing urban areas in each simulation period, so that if the specification is for the final service area to be 150 percent of the current urban areas at the start for five simulation periods, then the utility service areas are extended to 110 percent of the current urban areas at the end of the first simulation period and so forth. Note that any simulation zones within the current urban areas that were unserved at the start of the simulation will also be considered to be served by both water and sewer utilities at the end of the first simulation period.

The other option is to expand utilities as areas are developed, with the water and sewer utilities being extended to unserved simulation zones when the level of urban development (residential plus employment-related) exceeds the user-specified threshold. For implementations using regular systems of grid cells, such as the Central Indiana Implementation, such expansion will only take place if the unserved cell exceeding the threshold is adjacent to a cell that is already served. The process of extending utilities is repeated to allow expansion if an adjacent cell becomes served during that round. For implementations using irregular systems of simulation zones, the automatic inference of adjacency is not available, so utility service is provided through the automatic extension when the zone meets the urbanization threshold without consideration of service to adjacent zones.

Journey to Work

The journey to work model predicts the mean journey to work at the end of a simulation using the final forecast employment by employment zone and the final forecast population aggregated to the employment zones. Population values are adjusted so that the total of employment origins equals the total employment destinations.

The journey to work model described above is implemented within the simulation model and is run at the conclusion of each simulation. The accessibility coefficient used in the model is reduced from the estimated 2000 accessibility coefficient as described above. Using the employment origins and destinations, the distances between the employment zones, and the estimated reduced accessibility coefficient, the program calculates the balancing factors required by the double-constrained gravity model. The complete model is then used to predict the numbers of trips from each origin to each destination and to calculate the mean length of the journey to work.

Model Scenarios

This section describes the manner in which the various scenario options provided for in the model are implemented for the simulation.

Target Year

The target year specified determines the number of simulation rounds completed for the simulation.

Population Growth

The population growth for the region is calculated at the start of each simulation round using the option specified.

Density

During each simulation round, after the population density has been predicted by the model, the predicted densities are adjusted using the option specified. For the options to increase or decrease the density by a specified percentage, the predicted densities for the simulation zones (grid cells) are all uniformly increased or decreased accordingly. For the options for minimum or maximum densities of development, if the predicted density is less than the minimum (or greater than the maximum) it is replaced by the minimum (or maximum) density specified.

Utilities

A number of utility scenario options can be specified that are necessarily implemented in a variety of ways. If the options are selected to require water and/or sewer utility service for development, in each simulation round prior to the prediction of employment-related and residential development, the amount of land available for development in a simulation zone (grid cell) is temporarily set to zero if the zone is not currently serviced by the specified utility. (This is done on a temporary basis for each round, as subsequent utility expansion will result in changes in the areas served.) Likewise, for the option to require sewer utility service for development on land not suitable for septics, if the zone is not currently provided with sewer utility service and if the land in the grid cell is not suitable for septic systems, then the amount of land available for development is temporarily set to zero. (This option is not provided in the Central Indiana Implementation.)

For the option to expand utilities around major urban areas at a specified rate of expansion, the user specifies the final service area to be some percentage (greater than 100) of the areas of the urban areas at the start of the simulation. For the Central Indiana Implementation, the urban areas are associated with the urbanized areas and include those zones in which the percentage urban is greater than 25 percent (and other zones included within these areas). The simulation zones are ordered by distance from the urban areas and the model includes data on the cumulative percentage addition to the urban area for each simulation zone. At the end of each simulation round, utilities are expanded to include the proportional share of the total expansion for that round. For example, if the user specified that the final service areas should be 150 percent of the size of the initial urban areas for a simulation including 5 round, then expansion would be to 110 percent for the first round, 120 percent for the second round, and so forth.

For the option to expand utilities as areas are developed, a percent of land urban threshold must be exceeded for utilities to be provided to a simulation zone (grid cell). In the Central Indiana Implementation and any model using a system of regular grid cells, for expansion to occur (for a grid cell to be designated as being provided with utility services), the level of urbanization must have reached the threshold and the grid cell must be adjacent to a grid cell already provided with the specified utility service. For models using systems of irregular simulation zones, the latter criterion of service to an adjacent zone is not implemented because the necessary information will not be available. Areas are deemed to be provided with utility service once the threshold has been met without consideration of service provision in nearby areas.

Transportation

The implementation of the transportation scenario options depends upon whether the model uses distance or travel times in calculating the accessibility measures.

For a model using distances, such as the Central Indiana Implementation, the model begins with the distances from each simulation zone (grid cell) to the current transportation facilities (interstate highway interchanges and four-lane highways in the Central Indiana Implementation). These distances are updated at the beginning of each simulation round for any transportation options that have been specified for construction at that time. The model includes the distances from each zone to the various transportation alternatives included. For each zone for which the distance for the specified transportation option is less than the current distance for that transportation mode, the new, lower distance is substituted as the distance for that zone.

For a model using travel times, only one transportation option may be selected. For the year for which construction of the transportation option is specified, the new travel time matrix reflecting travel times with that transportation improvement is read into the model. The new travel time matrix replaces the original matrix and is then used for the calculation of all accessibility measures.

Sensitive Lands

The options to restrict development on specified sensitive lands are implemented at the start of the simulation. For any option specified, the quantity of the sensitive land in each simulation zone (grid cell) is subtracted from the amount of land available for development at the start of the simulation. If more than one type of sensitive land is specified for the restriction of development, the intersections of those types is added back to avoid double-counting the amount of land restricted.

Agricultural Land

The restriction of development on agricultural land is done in a manner analogous to that used for sensitive lands, with the further implementation of the additional options that may be specified. The option may be specified to restrict development on agricultural land only in those simulation zones (grid cells) that have at least some minimum percentage of land in agricultural use at the start of the simulation. So this check is made and any restrictions are applied only to those zones that meet this test.

The option may also be specified to restrict development on agricultural land only in those simulation zones that are at least some minimum distance from the edges of major urban areas. The model includes the distance from each zone to the nearest major urban area (defined as the areas associated with urbanized areas that are at least 25 percent urban at the start of the simulation). This check is made and the restrictions are applied only to those zones that meet this test.

The user may specify the percentage of agricultural land in a zone on which development is restricted. This percentage is multiplied by the amount of agricultural land in the zone at the start of the simulation to provide the amount of land that is subtracted from the initial amount of land available for development. If restrictions are also specified for any sensitive lands, the intersections with agricultural land restricted are added back to avoid double-counting.

Urban Growth Boundaries

For the option to establish urban growth boundaries around major urban areas, the user specifies the area within the urban growth boundaries to be some percentage (greater than 100) of the areas of the urban areas at the start of the simulation and the threshold level of beginning urbanization in the simulation zones that would allow development for zones outside of the urban growth boundaries. For the Central Indiana Implementation, the urban areas are associated with the urbanized areas and include those zones in which the percentage urban is greater than 25 percent (and other zones included within these areas). The simulation zones are ordered by distance from the urban areas and the model includes data on the cumulative percentage addition to the urban area for each simulation zone.

At the beginning of the simulation, those simulation zones that are outside of the urban growth boundaries are determined. Then, for those zones outside with levels of urbanization less than the threshold level, the amount of available land is set to zero, preventing any development from taking place. The amounts of available land are unchanged for zones inside the urban growth boundaries and for zones with starting levels of urbanization greater than the user-specified threshold.

Comprehensive Plan

Because the necessary data are not available, the option that development follow comprehensive plan guidelines is not included in the Central Indiana Implementation. When this option is included, data will be provided on the amounts of land in each simulation zone available for residential and employment-related development and the maximum or minimum density of residential development. The amounts of land available for residential and employment-related development by zone are separately identified in the model and used for the prediction of the respective forms of development. Checks are made to see that the total amounts of residential and employment-related development do not exceed the total land available for development in each zone at the start of the simulation after all restrictions on development have been implemented.

The maximum and minimum densities of development for each zone is implemented in the same manner, at the same time, as an overall maximum and minimum density of development specification is implemented, except that these maximum will generally be different for the various simulation zones.

The comprehensive plan option does not have to apply to all of the simulation zones and the specifications are only applied to zones for which specifications are provided in the file.

Dispersal of Development

This option allows the user to specify whether development should be more or less dispersed than otherwise predicted by the model. Both more and less dispersal are implemented by varying the predicted probabilities of residential development in the model, though in different ways.

The implementation of increased dispersal is accomplished by reducing the standard deviation of the distribution of the predicted probabilities of residential development. Maximum dispersal involves the reduction of the standard deviation to 90 percent of the standard deviation of the distribution of the probabilities as predicted. Lesser values of dispersal involve proportionately less reduction in the standard deviation.

Minimum dispersal is considered to occur when the simulation zones with the highest probabilities of development in each simulation round are totally developed and the remaining simulation zones have no development. So all zones are given probabilities of either one or zero. The number of zones to be given probabilities of one is determined iteratively so that development in those zones just accommodates the population growth. Intermediate levels of dispersal are implemented by interpolating between these minimum dispersal probabilities and the original predicted probabilities.

Accessibility to Employment

This option alters the relative importance of accessibility to employment for the location and density of new residential development. Different procedures are followed depending upon whether the user specifies increased or decreased relative importance for accessibility to employment.

For increased accessibility to employment, the accessibility to employment and accessibility to employment change coefficients for the model to predict the probability of development are increased. The maximum increase is set to a doubling of the accessibility coefficients. New accessibility values are calculated each round using these new coefficients, along with the original accessibilities. The new accessibilities are then standardized to have the same mean and standard deviation as the original accessibilities. (This is needed because the magnitudes of the accessibilities and the variations change dramatically with the coefficient. Using the unstandardized accessibilities calculated with the new coefficients produces counterintuitive results.)

The accessibility to employment coefficient for the model to predict the density of development is also increased by the same proportion. New accessibility values are calculated each round using the new coefficient, along with the original accessibilities. The new accessibilities are then standardized to have the same standard deviation and the same minimum value as the original accessibilities. (Standardization to the original minimum predicted accessibility is based upon the assumption that while the distribution of the accessibilities is changed, the minimum will remain constant. This was the method found that could appropriately set the level for the adjusted distribution.)

For decreased accessibility to employment, the accessibility to employment and accessibility to employment change coefficients for the model to predict the probability of development are decreased. The maximum decrease is set to a halving of the accessibility coefficients. New accessibility values are calculated each round using these new coefficients. (No standardization to the original mean and standard deviation are required for this case, because the reduction in the accessibility coefficient reduces the variation in the accessibilities, which is exactly what is desired.)

The accessibility to employment coefficient for the model to predict the density of development is also decreased by the same proportion. New accessibility values are calculated each round using the new coefficient, along with the original accessibilities. The new accessibilities are then standardized to have the same standard deviation and the same minimum value as the original accessibilities for those simulation zones in which development could occur (with positive probabilities of development and with positive amounts of land available for development). The final accessibilities used for the prediction of the density of development are then the lesser of the original accessibilities and the adjusted accessibilities. (The effect of the decrease in the accessibility coefficient on predicted densities and development made devising this procedure especially difficult. The lower coefficient increases both the magnitudes of the accessibility values and their variation. Standardizing using the values for zones in which development could not occur resulted in too great a variation in the final values. And by decreasing the variation, the standardization decreases the higher accessibilities but increases the lower accessibilities.)

For both increases and decreases in the relative importance of accessibility, the accessibility coefficient for the journey to work model is likewise increased or decreased. It is assumed that the increased importance of accessibility to employment will be reflected in the choice of locations that are closer to places of employment and will therefore affect the journey to work. Changing that coefficient by the same proportion resulted in unreasonable changes in the predicted mean journey to work. Since the changes in the journey to work would be associated with the new population living in new residential development, the proportional change in the accessibility coefficient for the journey to work model is the proportion of the final population that is population growth since the start of the simulation times the proportional change in the other accessibility coefficients.

Employment Growth

Employment growth for the employment zones (ZIP code areas) is calculated for each simulation round after population growth using the option specified. Details are provided in the section above on employment forecasting.

New Employment

After employment has been forecast for a simulation round, if new employment is specified to be added to a specified employment zone, that employment amount is added to the total employment for that zone and to the simulation round employment change for that zone and is used for all of the following simulation. For the cases in which service employment is being predicted, the additional employment is added to the basic employment for that zone. For the case in which employment is being predicted by industry, the new employment is added to the basic industries in proportion to their current employment.

Subarea (County) Growth

For the specification that the probabilities of development for a specified subarea (county) be increased or decreased by a specified percentage, the probabilities of residential development predicted in each round are adjusted up or down by the specified percentage for the simulation zones in the specified subarea. This adjustment is made before setting probabilities less than the minimum threshold to zero.

Other Variable (Education)

For the specification that differences in the other variable (education, ISTEP scores) be reduced, the standard deviation of the other variable is reduced as specified by the user. Complete reduction of differences (a selection of 10) reduces the standard deviation to zero, setting all values to the mean. Intermediate reduction of differences involves the proportional reduction of the standard deviation, i.e., a selection of 5 reduces the standard deviation by one-half.

User Scenarios

The user scenarios are implemented in ways analogous to the comparable scenarios in the model.

Restrict Development on Land Type

The amount of restricted land in each simulation zone is subtracted from the amount of available land in the zone after the implementation of any of the other restrictions of development on sensitive lands that may have been specified. If the resulting amount of available land becomes less than zero, the value is set to zero. This restriction is included in the calculation of the total amount of restricted land reported in the table.

Land Available for Residential Development

The amount of land available for residential development is implemented in the same manner as for the comprehensive plan option, setting the amount of available residential land to the value in the user land available for residential development file, but this cannot be more than the total amount of land available for development after restrictions. This can be applied only to specified simulation zones.

Minimum Density for Residential Development

This is implemented in the same way as the general minimum density specification and the minimum density specification for the comprehensive plan scenario. This can be applied only to specified simulation zones.

Maximum Density for Residential Development

This is implemented in the same way as the general maximum density specification and the maximum density specification for the comprehensive plan scenario. This can be applied only to specified simulation zones.

Land Available for Employment-Related Development

The amount of land available for employment-related development is implemented in the same manner as for the comprehensive plan option, setting the amount of available employment-related land to the value in the user land available for employment-related development file, but this cannot be more than the total amount of land available for development after restrictions. This can be applied only to specified simulation zones.

Water Service Areas

At the start of the simulation round for the year specified, if the simulation zone is specified as having water service in the user water-service file, the simulation zone is specified to have water service.

Sewer Service Areas

At the start of the simulation round for the year specified, if the simulation zone is specified as having sewer service in the user sewer-service file, the simulation zone is specified to have sewer service.

Transportation Alternative

If distances are being used in the model, at the start of the simulation round for the year specified, if the distance to the transportation alternative in the user transportation alternative file is less than the current distance to the specified transportation mode, the current distance is replaced by the distance from the transportation alternative file.

If travel times are being used in the model, at the start of the simulation round for the year specified, the specified user travel time matrix is read in and is used to place the updated travel times into the matrices for the calculation of the accessibilities.

Advanced Options

These are options that affect the basic operation of the simulation. Appropriate use of these scenario options requires an understanding of the model and the manner in which it simulates development.

Change Internal Demand for New Residential Development

As explained above in the description of the simulation of new residential development, the demand for such development is the sum of the forecast population increase for the region for the simulation period plus the demand for new development from existing residents. This internal demand is specified as the percentage of the current population demanding new development during a simulation period. The default value included in the Central Indiana Implementation was estimated for the period from 1990 to 2000. This option allows the user to change this percentage for the internal demand for new residential development.

This scenario option can be used as a crude proxy for policies that would encourage residents in existing residential areas to remain in those areas, perhaps through various types of neighborhood improvement programs.

If the option to use county control totals is selected, no internal demand for new residential development is included.

If the user wants to output simulation results and use the simulated populations for the simulation zones, the value of the internal demand should be set to zero. A positive internal demand value will result in additional population being allocated with new residential development without the concomitant reduction in populations in existing residential areas. As a result, with a positive value for internal demand, the total of the simulated population for the simulation zones will be greater than the total forecast population for the region.

Change Maximum Amount of Land Available for Urban Development

Historical results suggest that in newly-urbanizing areas, the amount of new urban development seldom reaches 100 percent of the available land. This option allows one to set the maximum level of such development. It is implemented by setting the amount of available land in each simulation zone to a value so that if that land were fully-developed, the amount of urban development in the zone would be equal to the specified percentage.

Change Minimum Probability for Residential Development

As explained above, all of the probabilities of residential development predicted by the model are greater than zero because of the nature of the logit model. Using the smallest probabilities in the allocation of residential development results in the unrealistic spreading of residential development to all simulation zones in the area. A minimum probability is set for residential development, and predicted probabilities less than this value are set to zero prior to the allocation of residential development. This option allows one to change that minimum probability.

Change Criteria for Allocation of Employment-Related Land Use

As explained above in the description of the simulation of employment-related development, a minimum value is set for the amount of employment-related development to be allocated to a simulation zone and this development will in general be allocated to a specified maximum percentage of the simulation zones within an employment zone. These options allow one to vary those parameters.

Model Implementation

The *luci2 Urban Simulation Model* was developed during the period 2003 to 2004. Most of 2003 was spent on creating the land use data and estimating the various models as described above. Development of the program for the model began in the latter part of 2003. The model was fully functional by August, 2004, with the first release of the model for beta testing occurring October 28, 2004. All of the model development, including all of the coding of the program, was done by John Ottensmann

Model Development

The program was developed with Microsoft Visual Basic 6.0, using only the capabilities that are provided with the program. No additional externally-developed controls are used in the program. (Visual Basic 6.0 was used rather than the newer Visual Basic.Net, as the graphics commands in the earlier version made writing the mapping routines significantly easier.)

The program uses the Visual Basic Multiple Document Interface. Virtually all of the program elements (command buttons, the picture box for the map, the grid control for the tables, the controls for setting the scenario options, and so forth) are on child forms that are displayed within the main form. The exceptions are the dialog boxes for saving and loading scenarios and for printing, which use standard Windows dialogs, message boxes that appear for warning, and the Help form, which is displayed as a separate window.

All of the data required to create a scenario are loaded at program initialization and are stored in memory. This is to provide for maximum speed in running the simulation. For scenarios specifying the use of an external employment forecast file or for the use of a new travel time matrix for transportation options (when travel times are used), these data are read in during the simulation.

The Help window uses the simple Visual Basic web browser to display the help information. All of the help files are standard *html* files. These are stored in the *Help* folder under the *luci2* program folder.

All of the program data files and setup files and the scenario files and program output files generated by the program are simple, standard comma-separated variable files. The program data files, setup files, and program output files are stored in the *Data*, *SetupFiles*, and *ModelResults* folders under the *luci2* program folder. Scenario files provided with the model are in the *Scenarios* folder, which is the initial default location for saving new scenario files. However, scenario files may be saved to and loaded from any folder using the standard Windows dialogs. The scenario files have the extension *.luc*. All of the other files have the extension *.txt*.

All data required for the model were organized and saved in an Access database. The data files required by the model were produced by exporting appropriate tables from that database. The required and optional data files used by the model are documented in *luci2DataFiles.doc*.

Implementation Specification

As described earlier in this document, the *luci2 Urban Simulation Model* is a general-purpose urban simulation model program that can be used to implement models of varying complexity, with varying features, for different areas. The model definition file, *ModelDef.txt*, provides all of the settings and parameter values required for a specific implementation. It is a long file, with over 400 values associated with defining and implementing a specific implementation. This file and all of the settings contained therein are documented in *luci2ModelDefinitionFile.doc*.

The *ModelDef.txt* file is the first file read at program initialization. Settings in that file determine which data files are required for the specific implementation and control which data files are read at initialization. These settings also determine which elements are included in the user interface and, in some cases, captions or labels associated with those elements. For example, the scenario option to require that development follow comprehensive plan requirements is not included in the *Central Indiana Implementation* (because appropriate data were not available). Therefore, that button does not appear with the other *Create Scenario* buttons. The settings also determine which maps and tables are included and what information is to be included within the tables.

Model Development Tools

For the estimation of the models contained within *luci2*, as described in an earlier section of this document, two small programs have been developed as model development tools, the *luci2 Accessibility Tool* and the *luci2 Journey to Work Tool*.

Estimation of the models requires value for accessibility, either to employment or population (or changes in those). These accessibility values are then included as independent variables in the regressions used to estimate the model. The *luci2 Accessibility Tool* is a small program used to calculate those accessibilities. The user specifies the accessibility value to be created, the value being predicted in the regression model. Then the user has the option of either having the program find the value of the accessibility coefficient that results in the best prediction of the value being predicted or the user can enter any value for the accessibility coefficient. The user can then save the accessibility values to a file which can then be read into the statistical program being used for the regressions.

Since the regression models being estimated have other predictors in addition to accessibility, the accessibility coefficient that results in the best prediction in the *Accessibility Tool* will not in general result in the best prediction in the regression model. So the strategy was to first use the *Accessibility Tool* to find the value for the accessibility coefficient that provided the best (univariate) prediction and then also creates accessibility values for coefficients above and below that. These accessibilities reflecting the different accessibility coefficients were then each tested in the regression model to find the range for the accessibility coefficient within which the best prediction fell. Then the *Accessibility Tool* was used to produce additional accessibility values for coefficients within that range. This process was repeated until the accessibility coefficient resulting in the best prediction was established to two significant digits.

The accessibility coefficient for the journey to work model and the change in that coefficient over time were estimated using the *luci2 Journey to Work Tool*. This program allow the user to specify a value for the accessibility coefficient, calculates the predicted trips, the mean predicted journey to work, and various measures of the goodness of fit of the predicted trips to available data. The measures and the use of the model are described more fully in the section on the estimation of the journey to work model.

Enhanced User Capabilities

As described earlier in this document, the model included several enhanced user capabilities, the ability to create new user-specified scenarios and the option to automatically output extended model results during a simulation. Whether the user of the model will have access to these additional capabilities is specified in the model definition file.

User-specified scenarios are defined in the user scenario definition file, *UserScenarioDef.txt*. This is documented in the file *luci2UserScenarioDefinitionFile.doc*. The settings in this file specify which user scenarios are to be provided and the descriptions to be used for these scenarios in the forms and tables. They determine which additional data files need to be read in at program initialization for the user scenarios. The data files required for the user scenarios are described along with all of the other data files in *luci2DataFiles.doc*.

The enhanced capabilities also provide the user with the ability to specify that model results—land use and population for the simulation zones and employment for the employment zones—be automatically saved when a scenario is being simulated. The user has the option of no results, values for the start and end of the simulation or, additionally, values for the end of each intermediate simulation round. These results are saved as comma-separated variable files with variable names on the first line in the *ModelResults* folder. Two file is saved for each year, one for the simulation zones and one for the employment zones. The filenames have the prefix "*Sim*" or "*Emp*", the name that the user provided for the scenario, and the year for the results, with the extension *.txt*.

Command Line Version

A separate version of the model will be developed that can be run from the command line that will run a simulation and output the simulation results without displaying any forms, requiring no user intervention. This command line version is intended for applications in which the *luci2 Urban Simulation Model* is being integrated into a system with other models operating under a common user interface. The command line version could also be employed in the implementation of a web-based version of the model that might or might not involve integration with other models.

The command line version of the model will employ a scenario definition file that includes all of the scenario option values for the scenario to be generated. The filename for this file will be included on the command line. The program will read this file, generate the specified scenario, and output the simulation results.

Developers of systems using the command line version will have the obligation of providing the scenario definition file for the desired scenario and, obviously, implementing whatever they intend to implement with the simulation results. The developer could create a user interface to obtain the scenario options from the end user of the system and then use that information to create the scenario definition file. (Such an interface could implement as many or as few of the *luci2* scenario options as desired.) Alternatively, the developer could create a set of scenario definition files and allow the user to select from this set.

The command line version will write the simulation results to files in the same manner as when a user of the regular version with the enhanced user capabilities chooses to automatically output results. In addition, the command line version will output a results file at completion. This will include a code indicating whether the simulation was successfully completed or whether it failed because of problems in reading the scenario definition file, or whether the simulation failed for some other reason. It will also include a setting indicating whether the simulation resulted in running out of land, an event that would not be included in the model results files.

Model Distribution

As with the original LUCI model, the Central Indiana Implementation of the *luci2 Urban Simulation Model* will be freely distributed and will be made available on the LUCI website at http://luci.urbancenter.iupui.edu/. This freely distributed version of the model will not include the capabilities for creating new implementations of the model and will not include the enhanced user capabilities. In all other respects, however, this version will have all of the other model capabilities.

Rights to use the *luci2 Urban Simulation Model* to create new implementations and rights to use the enhanced user capabilities will be retained by John Ottensmann and the Center for Urban Policy and the Environment. These capabilities will not be freely distributed.

Protection of these rights to use the model will be implemented in two ways. First, at a very basic level, the files documenting the model files, the model definition file, and the user scenario definition file will not be made generally available. Creation of a new implementation would be extremely difficult (though not impossible) without this information. (The model definition file, for practical reasons, is reasonably self-documented. Figuring out the structure of the data files would be more difficult. While the data are grouped into multiple files with, again for practical reasons, descriptive names, the variables contained in each file are not identified within the files.)

In addition, further protection of these rights is provided by the inclusion of a security key in the model definition file. This key is generated by using some of the basic information required for initially defining the model that is included in the model definition file. At program startup, the key value is computed from the information in the model definition file and compared with the security key in the file. Modification of that basic information will result in the key values not matching. The user will be notified and the program will terminate.

It may be desirable to provide certain users of the *Central Indiana Implementation* with access to the enhanced user capabilities without giving them the ability to implement new models. This is accomplished by providing for a time-limited key allowing the use of the enhanced capabilities until a certain date. The key will be provided in the form of a very brief text file with the end date and a security key based upon that date. At startup, the program will look for that file. If found, it will read the file, compare the key in the file to the key value generated using the file date (to prevent user editing of the date), and then compare that date to the system date to determine whether the enhanced capabilities are to be provided.

References

Berkson, J. 1953. "A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function." *Journal of the American Statistical Association*, 50: 130-62.

Berkson, J. 1955. "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function." *Journal of the American Statistical Association* 48: 565-99.

Cox, D. R. 1970. *The Analysis of Binary Data*. London: Methuen.

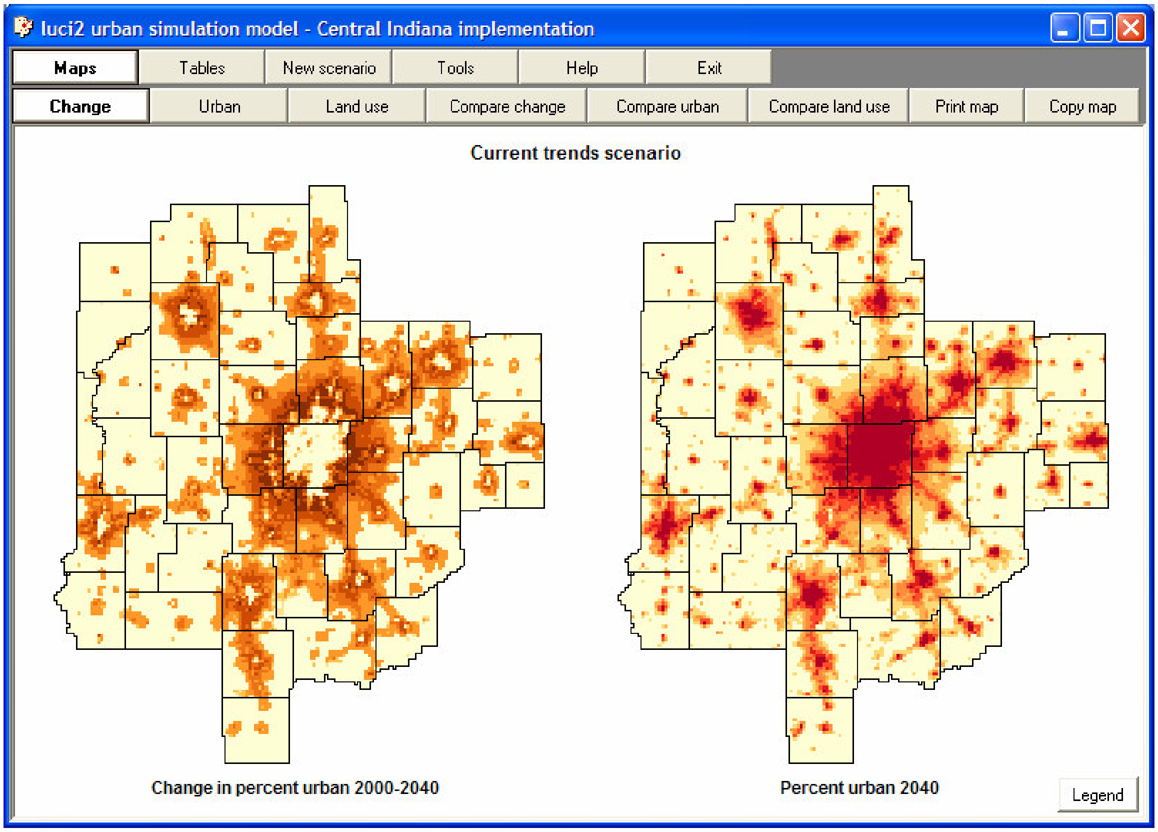
Morisette, Jeffrey T., Cheshire, Heather, Stalling, Casson, & Khorram, Siamak. 1996. "An Urban Mask Raster Image for Vector Street Files." In Stan Morain and Shirley Lopez Baros (eds.), *Raster Imagery in Geographic Information Systems*, pp.172-178. Santa Fe, NM: OnWord Press.

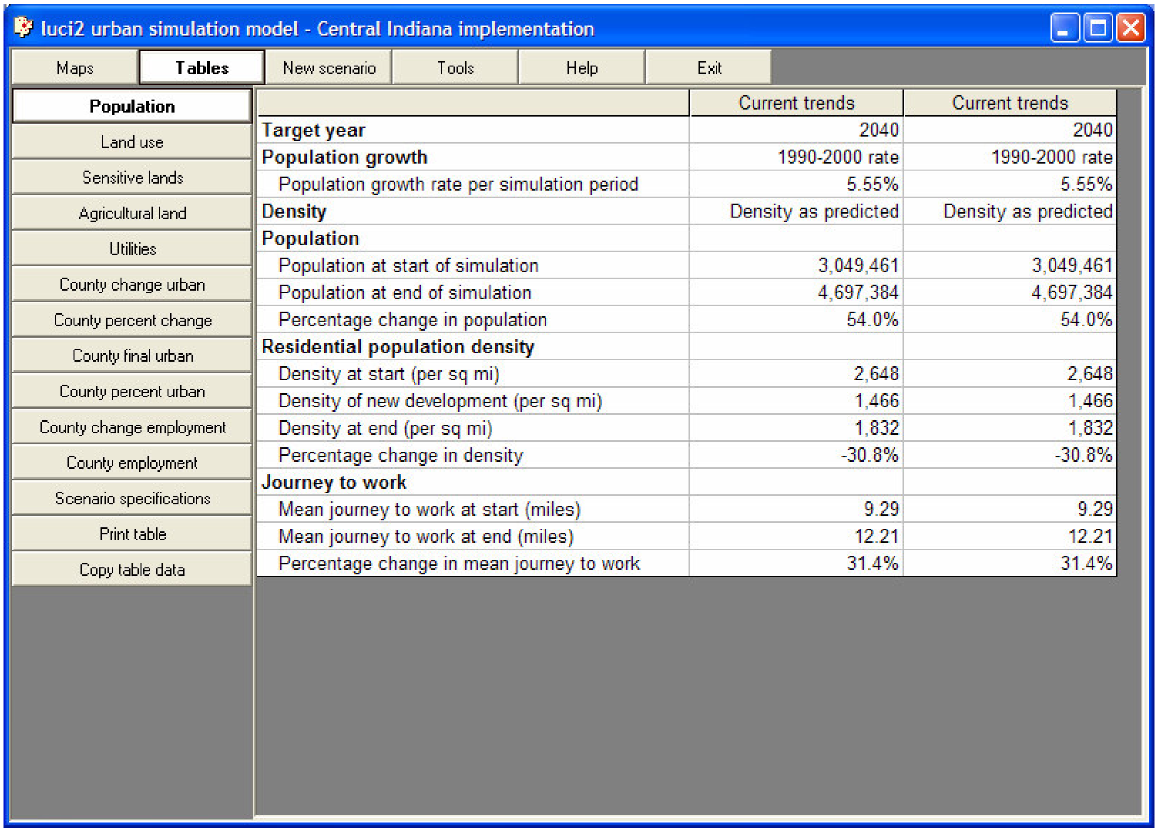
U.S. Bureau of the Census. 1998. *ZIP Code Business Patterns, 1995 on CD-ROM*. Prepared by Microdata Access Branch, Administrative and Customer Services Division, Bureau of the Census. Washington: The Bureau, 1998.

U.S. Bureau of the Census. 2003. *ZIP Code Business Patterns, 2000*. Downloaded from http://censtats.census.gov/, June 2003.

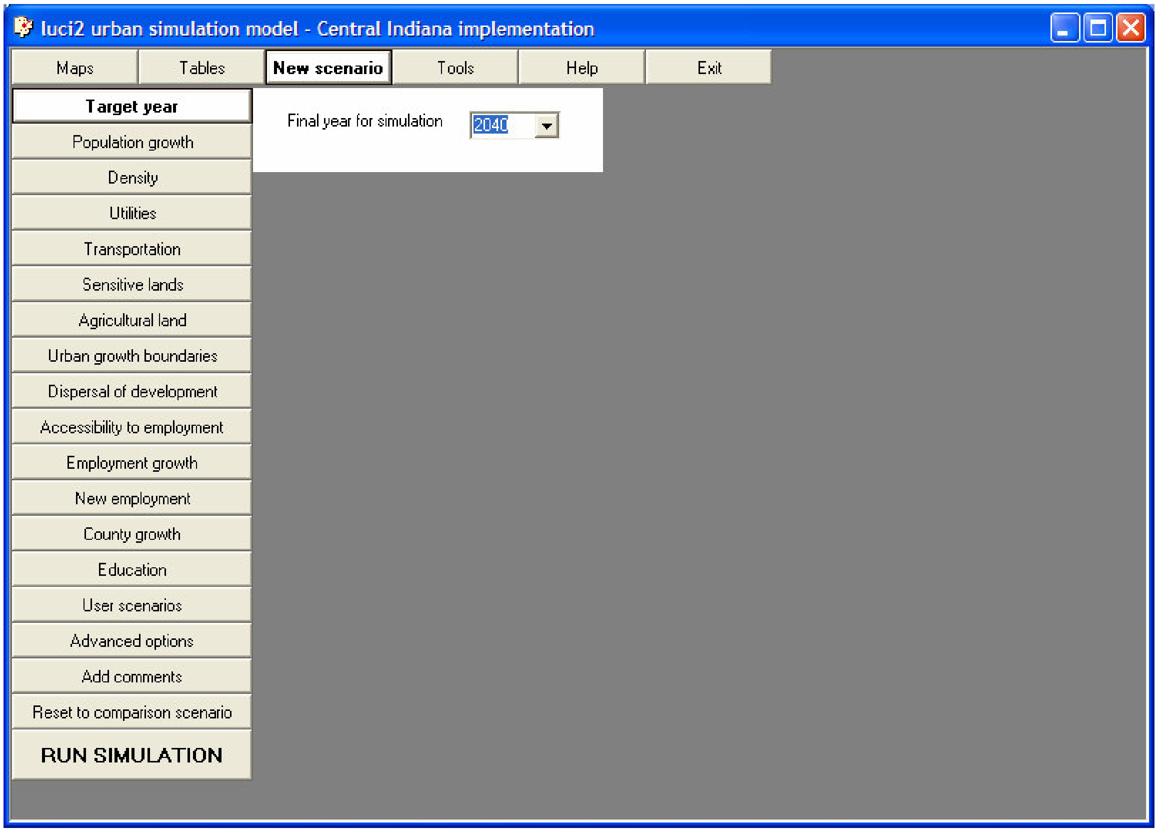
Wrigley, Neil. 1985. *Categorical Data Analysis for Geographers and Environmental Scientists*. New York: Longman.

Appendix
luci2 User Interface

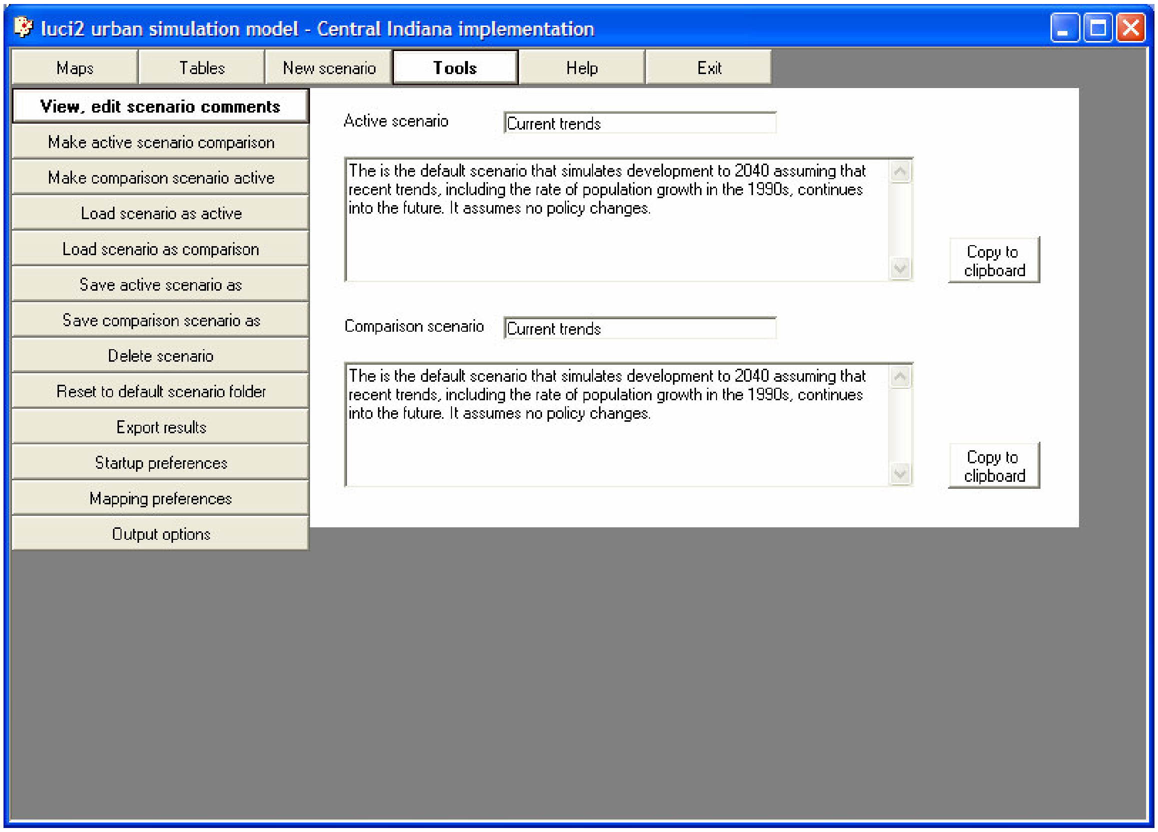




	Current trends	Current trends
Target year	2040	2040
Population growth	1990-2000 rate	1990-2000 rate
Population growth rate per simulation period	5.55%	5.55%
Density	Density as predicted	Density as predicted
Population		
Population at start of simulation	3,049,461	3,049,461
Population at end of simulation	4,697,384	4,697,384
Percentage change in population	54.0%	54.0%
Residential population density		
Density at start (per sq mi)	2,648	2,648
Density of new development (per sq mi)	1,466	1,466
Density at end (per sq mi)	1,832	1,832
Percentage change in density	-30.8%	-30.8%
Journey to work		
Mean journey to work at start (miles)	9.29	9.29
Mean journey to work at end (miles)	12.21	12.21
Percentage change in mean journey to work	31.4%	31.4%



Final year for simulation 2040



Active scenario: Current trends

The is the default scenario that simulates development to 2040 assuming that recent trends, including the rate of population growth in the 1990s, continues into the future. It assumes no policy changes.

Comparison scenario: Current trends

The is the default scenario that simulates development to 2040 assuming that recent trends, including the rate of population growth in the 1990s, continues into the future. It assumes no policy changes.